Violin Augmentation Techniques for Learning Assistance

Laurel S. Pardue

PhD Thesis

School of Electronic Engineering
and Computer Science
Queen Mary University of London

Submitted in partial fulfilment of the requirements of
the University of London Degree of Doctor of Philosophy

2016

I, Laurel Smith Pardue, confirm that the research included within this thesis is my own work or that where it has been carried out in collaboration with, or supported by others, that this is duly acknowledged below and my contribution indicated. Previously published material is also acknowledged below.

I attest that I have exercised reasonable care to ensure that the work is original, and does not to the best of my knowledge break any UK law, infringe any third party?s copyright or other Intellectual Property Right, or contain any confidential material.

I accept that the College has the right to use plagiarism detection software to check the electronic version of the thesis.

I confirm that this thesis has not been previously submitted for the award of a degree by this or any other university.

Signature:

// Laurel S. Pardue //

Date: 16, Nov 2016

Details of collaboration and publications:

Laurel S Pardue, Christopher Harte, and Andrew P McPherson. A low-cost real- time tracking system for violin. *Journal of New Music Research*, 44(4):305–323, 2015.

Laurel S Pardue and Andrew P McPherson. Near-field optical reflective sensing for bow tracking. In *Proceedings of the 2013 Conference on New Interfaces for Musical Expression*, 2013.

Laurel S Pardue, Dongjuan Nian, Chris Harte, and Andrew P McPherson. Low- latency audio pitch tracking: a multi-modal sensor-assisted approach. In *Proceedings of the 2014 Conference on New Interfaces for Musical Expression*, 2014.

Abstract

Learning violin is a challenging task requiring execution of pitch tasks with the left hand using a strong aural feedback loop for correctly adjusting pitch, concurrent with the right hand moving a bow precisely with correct pressure across strings. Real-time technological assistance can help a student gain feedback and understanding helpful for learning and maintaining motivation. This thesis presents real-time low-cost low-latency violin augmentations that can be used to assist learning the violin along with other real-time performance tasks.

To capture bow performance, we demonstrate a new means of bow tracking by measuring bow hair deflection from the bow hair being pressed against the string. Using near-field optical sensors placed along the bow we are able to estimate bow position and pressure through linear regression from training samples. For left hand pitch tracking, we introduce low cost means for tracking finger position and illustrate the combination of sensed results with audio processing to achieve high accuracy low-latency pitch tracking. We subsequently verify our new tracking methods' effectiveness and usefulness demonstrating low-latency note onset detection and control of real-time performance visuals.

To help tackle the challenge of intonation, we used our pitch estimation to develop low latency pitch correction. Using expert performers, we verified that fully correcting pitch is not only disconcerting but breaks a violinist's learned pitch feedback loop resulting in worse as-played performance. However, partial pitch correction, though also linked to worse as-played performance, did not lead to a significantly negative experience confirming its potential for use to temporarily reduce barriers to success. Subsequently, in a study with beginners, we verified that when the pitch feedback loop is underdeveloped, automatic pitch correction did not significantly hinder performance, but offered an enjoyable low-pitch error experience and that providing an automatic target guide pitch was helpful in correcting performed pitch error.

Acknowledgements

First, a huge thanks to my primary supervisor Dr. Andrew McPherson for his support, insight, and overall being brilliant. Every meeting we had I expected to come out with better ideas and more clarity. I couldn't ask for a better advisor.

I'd also like to thank Dr. Elaine Chew and Dr. Tony Stockman for their help shaping thesis content and direction; Kok Ho Huen, David Wilkinson, Geetha Bommireddy for their help when I needed things in the electronics lab, and Melissa Yeo for admin support.

A huge thanks to both Kate Conway and Sigurd Feiring for their support in the study with beginners, taking the risk of letting me work with their students, and making substantial efforts to help a stranger. Thanks to Dave Moffat for lending me his headphones for ages to use in user studies; Luwei Yang for sharing his work on vibrato detection; Nan-Wei Gong for help getting me started with conductive printing; Liz Routhorn for volunteering to proof this document and doing a fantastic job, and Siddarth Sigita for all the various favors over the years. Big thanks to my excellent research group, the Augmented Instruments Lab and the many many others around C4DM whom have made my time at QMUL fun, interesting, and inspirational.

Lastly thanks to Bex Burch for feedback and belief, my flatmates Fran and Kate for support and tolerance, and all the many friends whom have contributed ideas, good will, or good conversation over the course of this thesis.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Motivation and Aims

Violin is a much loved instrument. It has a rich history featuring hundreds of years of virtuosic performance as a popular solo instrument including hundreds of classical concertos and a firm place in many folk traditions. The violin is typically considered one of the most expressive instruments but also one of the most difficult. Unlike most instruments, such as piano, guitar, and saxophone, pitch on the violin is continuous with no frets, keys, or other obvious visual or tactile guides as to where a note is and what is correct. If accuracy of pitch, amongst near infinite options were not challenging enough, sound on the violin is not created through direct human action, but through control of an external implement, the bow.

Violinists, and players of other bowed string instruments, must deal with the challenge of playing in tune. Playing in tune requires performers to develop precise motor memory to locate notes and an accurate fast ear to evaluate whether their initial estimate is correct. But while the challenge of finding the right note is more obvious to the uninitiated, just as difficult is learning to use the bow.

It takes on average 700 hours just to learn basic bow technique [89] and far more before it is mastered. Mastering bow technique requires negotiating control of seven different parameters such as pressure, tilt, skew, and bow-bridge distance [151]. As the renowned violin virtuoso

Tasmin Little stated during a debate on instrument difficulty[1], "...you can be doing everything right on the violin, but as 90% of tonal production comes from the bow, as long as your bow isn't working, nothing is going to work."

What further complicates violin learning is that neither intonation nor bowing are easy to simplify. Attempts to focus on bowing or pitch only are deeply unsatisfying. Using only the bow a musician can only play four notes, the set of which is not particularly useful for melody. Focusing exclusively on pitch, the only alternative to the bow is plucking the strings which does not produce enough sustained sound to accurately evaluate intonation. The necessity to be able to do everything to achieve anything yields a perfect storm of frustration and unenjoyable performance.

Considering both the violin's popularity, and the high level of difficulty in learning, it is an obvious candidate for the application of digital technology to assist with learning. Although there have been some investigations into technologically-assisted violin learning aids, such as i-Maestro [120] and Schoonderwaldt's explorations [157], both video motion capture based, and Musicjacket [169] which uses a motion capture suit (all discussed in depth along with several more in Section 3.3), they largely use highly expensive or complex technologies making it infeasible to deploy them for the average student's home practice. In fact, due to the complexities and expense of technologies involved, these tools have been predominantly restricted to laboratory studies with research completed in an academic environment.

So far, practical technologically-assisted violin learn aids have been confined to video-conferenced lessons, a technologically-assisted version of traditional teaching, and non-interactive instructional videos. Only Johnson [76], focusing primarily on vibro-tactile feedback for targeted help with specific bow and posture tasks, has attempted to design practical distributable tools for violin learning and has also tested these in broader real-world scenarios. Despite the huge market of potential violin students, there are no commercial options.

With such nuanced control in both hands, it is not surprising that it is challenging to track all the performance parameters needed to provide useful pedagogical feedback. Much of the restriction of learning tools to laboratory contexts is due to the difficulty of developing

[1] *Violin or Guitar: Which is Easier?* Originally broadcast on 16 Sept 2014 for BBC Radio 4 Today show with guitarist John Etheridge.

effective, deployable violin tracking mechanisms. Controlling pitch and bow require such a highly developed kinaesthesia that it is easy for sensors to physically interfere with normal play and mounting them on delicate expensive instruments can be problematic.

Considering both the lack of potential real-world applicable violin learning aids and the challenges of making music on the violin as a beginner, this thesis targets practical practice tools to make violin learning easier and to make the learning process enjoyable earlier. We ask, 1) can technology be used to reduce the impact of poor bow technique or poor pitch accomplishment, making the instrument more accessible and rewarding at all levels? And 2) can technology aid a major practice task like learning intonation?[2]

With these questions in mind, in this thesis we start by presenting an augmented violin: a violin whose capabilities have been extended through sensors and audio processing, for both pitch and bow tracking. Our augmented violin is designed with the purpose of being useable for learning in a wide range of contexts outside the laboratory, to enable both study of the practice aid in real-world situations and eventual wider use. We subsequently demonstrate our augmented violin's general usability and potential usefulness in helping students learn pitch in a real-world teaching context, and lastly, use it to investigate ideas for artificially simplifying violin playing in the pursuit of letting a beginner struggling with the difficulty and complexity of the instrument 'just play.'

1.2 Research Themes

Research in this thesis is divided into two main themes. Firstly, with the current lack of a suitable cost-effective augmented violin for use outside the laboratory, we start with design and implementation of a low-cost real-time augmented violin, with separate approaches to left-hand tracking and bow tracking. With the development of core tracking technologies, we explore and verify our augmented violin's capabilities. Our second theme is the investigation of approaches to aid violin practice and learning.

In the context of these two themes, we take a moment to consider both more specifically.

[2] As will be discussed in more depth in Section 2.4.3, for the purposes of this thesis, we are focusing exclusively on the equal temperament tunings used in Western Music

1.2.1 A General Use Augmented Violin

Beyond practice aids, violin tracking enables exciting opportunities focusing on *how* one plays, not just *what* one plays. Applications of the technology include study of stylistic differences between strokes [144] or traditions [156], study of ensemble cohesion [62], performance with automated systems [20, 8], and expansion of performance capabilities [110]. Many specialized systems (both real-time and non-realtime) [2, 184, 8, 37, 155] have been designed to pursue these worthwhile targets. However, the goal of a low-cost violin tracking system useful in practice has yet to be reliably accomplished.

Augmented Violin Design Objectives

Obvious requirements for a general real-world-use augmented violin, as supported by Johnson [76, p.87-88], are low cost, robustness, portability, and ease of use. Considering much of basic technique in violin performance is developed through proprioception, the kinaesthetic sense of the relative positions of the body, real-time feedback is more helpful in beginner studies than reflective learning [85]. Further, research has found that motor performance in music is significantly disrupted if auditory feedback is delayed [190], meaning any audio intervention must be real-time. Violins are also expensive and usually highly personal, meaning that, for any expectation of sustained use, augmentations should be useable on the performer's own instrument but must not damage it. Ideally, augmentations should be reversible. Lastly, any augmentation for the purpose of learning is only useful if it supports good technique. Augmentations must be minimally intrusive and not negatively interfere with normal play.

The augmented violin introduced in this thesis is a low-cost real-time system for both pitch and bow tracking. We consider low-cost to be below $150 USD and aim for the accepted target of sub-10ms latency for real-time interactive audio systems [49]. The system is designed to fit any violin and bow with easy installation, minimal impact to normal playing, and no damage to the instrument. With our focus on generalized use and learning, our augmented violin is not intended for the high accuracy required in musicological and acoustical studies such as [24] and [26].

Our augmented violin consists of the following two differing sets of technology:

List 1 Design objectives for a general use augmented violin:

1. Low cost
2. Real-time
3. Accurate
4. Robust
5. Portable
6. Easy for general public use
7. Minimally intrusive on normal play
8. Non-destructive and useable on any instrument
9. Supports traditional practice methods

1. Bow tracking of position and pressure using optical sensors (Chapter 4).
2. Pitch tracking using sensor-assisted pitch estimation for low latency and high accuracy (Chapter 5).

1.2.2 Assisting and Simplifying Violin Learning

Previously we mentioned learning aids based on motion capture video systems and suits but most violin learning occurs during individual practice. Lessons show students how to play. but as practice occurs mostly in the home without knowledgeable help available, it is the environment most important for engaging beginners. Any technological intervention must be both functional and understandable outside the academic environment. Although Johnson [76] has developed useful tools for practice and tested them 'in the wild', Johnson's real-world aids have very limited application, focusing on specific posture and bowing tasks. The limitations of Johnson's teaching aids are to some degree intentional in order for interactions to remain task-focused and simple, but it remains valuable to produce an aid with broader utility tackling a more generally problematic task like intonation.

Learning and Practice Aid Research Objectives

We want to consider a broader view of technological intervention in practice to both help accomplish major practice tasks, and encourage long-term practice motivation so a student, young or old, makes substantial strides towards completing the 10,000 hours of practice necessary for mastery [41]. Beyond the basic objective,

to explore the use of technology to assist violin learning,

we looked at two ideas for assisting learning, one based more on traditional violin pedagogy and one more radical.

The first idea focuses on assisting intonation. We use two approaches, the first is inspired by the common teaching practice of playing with a student. By playing with the student, the teacher provides an example of the piece played correctly that the student can follow and compare themselves to. The second approach is inspired by modern instrument tuning using visual cues from a digital tuner. As research questions, we ask:

1. Does providing the beginner with an aural guide of correct pitch help improve a student's pitch accuracy?
2. Does providing the beginner with visual feedback for correcting pitch help improve a student's pitch accuracy?
3. Does combining modalities, using both an aural guide and visual feedback help improve a student's pitch accuracy?

The second idea focuses on extended practice engagement. Moving away from traditional pedagogy, we investigate the use of technology to alter how the instrument plays based on *complexity management*. The idea of complexity management originates from a combination of traditional pedagogy and Csikszentmihalyi's [32] concept of 'game flow', where the difficulty of a game is ideally matched to the player's skill. Starting from Jorda's idea of music instrument efficiency [77], that looks at effort versus skill in musical instruments, we suggest factoring in musical reward, a major motivational factor in instrument play, for a more refined alternative to the learning curve: *learning efficiency*. We are interested:

to study how we can improve learning efficiency through the management of instrument complexity.

Though borne out of consideration of learning in new digital instruments, the violin is a perfect example of a high-complexity instrument with unsatisfactory early musical rewards. The violin's poor learning efficiency makes it an excellent instrument with which to test our ideas on complexity management. In order to do this, we examine how to improve the balance between user effort to reward using the augmented violin to make a violin easier. We look at how to artificially relax the exacting demands from the bow or from pitch execution instead offering alterable difficulty levels better matched to the user's skill.

The idea behind complexity management is not necessarily to improve learning in the short term, but to maintain instrument motivation in the longer term. As will be discussed in Section 2.3, full, extended research into complexity management is beyond the scope of this thesis, but this thesis constitutes a beginning. Within this thesis, we propose both complexity management as a useful idea and look at the viability of applying it to the violin including where it might yield potential benefits in real-world learning situations. We start with initial investigations into simplifying pitch by using the augmented violin to replace performer's actual performance with one variably corrected for pitch. We then ask how this impacts actual performance, user experience, and whether it offers learning opportunity and motivation.

1.2.3 Contributions

This thesis and research objectives within our two research themes are expected to contribute in a number of ways, technically, pedagogically, and theoretically. Chapters 2 and 3 provide overall context and background of the work, and the subsequent chapters present the contributions themselves.

1. Design and use of optical sensors for low-cost, low-latency, real-time bow tracking (Chapter 4).

2. Low-latency, real-time pitch and note estimation for the violin through fusion of sensor and audio analysis (Chapter 5).

3. Through the bow tracking and finger tracking we obtain novel means for detecting note onset through off-string to on-string transitions and fingering changes (Chapter 7).
4. Demonstrated effects of altering levels of pitch feedback on violin performance and the link between user skill level and perceived experience (Chapter 8, Chapter 9).
5. Demonstrated potential to improve pitch performance through providing an in-tune aural guide and high speed visual feedback for performance conducted through a real-world study with beginners (Chapter 9).

1.3 Thesis Outline

This thesis is comprised of nine chapters structured as follows:

Chapter 2 looks at how a student learns violin and intonation and introduces complexity management. We start with a brief introduction to violin terminology and basic performance technique before moving onto pedagogical aspects of the learning process and practice. Discussion then proceeds to the effects of an instrument's learning curve and Jorda's concept of musical efficiency. We suggest instrument learning efficiency as an extension of the learning curve that explicitly reflects how musical reward is key for practice motivation and formally introduce the idea of complexity management for improving an instrument's default learning efficiency. Finally we discuss the importance of and process for good intonation, including ideas for using pitch snap, a form of complexity management, to simplify the requirements of correct intonation.

Chapter 3 presents prior work relevant to the ideas in this thesis. It includes a review of violin augmentations, to both bow and fingerboard, and how those augmentations have been used. We include a review of selected music practice technologies with an in-depth review of technologies applicable to violin practice. The chapter also reviews current pitch estimation algorithms, with an emphasis on algorithms effective for violins, or that use sensor and audio information combined.

Chapter 4 describes a method for bow tracking using optical sensors. It introduces what we call the *displacement triangle* as a measurable means of uniquely describing most bow position and pressure combinations. The chapter then reviews the practical use of analog and digital optical sensors including circuitry before introducing polynomial fitting for estimating bow position. With the basic concept for position estimation presented we add tracking bow pressure as well. We include results from tests proving performance sufficient for pedagogical feedback.

Chapter 5 discusses a method for improving low-latency, real-time pitch estimation on the violin. The chapter proposes using a custom fingerboard sensor to estimate approximate pitch which is then used to inform traditional pitch estimation means for higher accuracy results. This reduces large pitch error and enables estimation of low frequencies within a short time period for an overall increase in high-speed pitch estimation accuracy. The chapter includes all information on sensor build, circuitry and algorithms needed for sensor-assisted pitch estimation, along with a comparison of performance with non-sensor assisted pitch estimation methods in a test using an acoustic violin. This chapter also introduces the two primary augmented violin designs, one electric and one acoustic, used in subsequent chapters.

Chapter 6 presents a case study of using the augmented violin to identify note onset in real-time, a difficult task with stringed instruments. Employing sensor only techniques we demonstrated the augmented violin was capable of matching or beating existing state-of-the-art real-time audio only detection techniques in 4 out of 5 types of note onset. Using position and pressure estimates from the augmented bow we were able to identify note onsets produced by off-string attack, and slurred repetition in real time. The fingerboard sensor also allowed us to detect note onsets produced by changes in fingering in real-time. Bow changes were not effectively detectable in real-time but become distinguishable with the real-time constraint removed.

Chapter 7 focuses on applications of the augmented violin including implementation details and demonstrating its use with pitch correction and outside a learning context. The chapter describes a software VST implementation for the augmented violin combining both bow tracking and pitch tracking together and sending results onward for further use. A pitch-retuning VST used in our studies is introduced that uses the low-latency pitch estimates from the augmented violin VST for retuning violin audio. Finally, we detail use of the augmented violin in live audio-visual performance.

Chapter 8 describes the initial pilot study with the electric augmented violin using expert violinists. The study aims to verify acceptability of the augmented violin and the pitch retuning VST, and to accomplish a viability test of the ideas behind complexity management when applied to performing pitch. Participants wore headphones to block acoustic audio, instead hearing variably pitch snapped audio, in order to test the effects of increasing heard pitch accuracy while reducing ability to assess error. The chapter discusses study design, execution and results including a discussion of user reaction to the instrument itself and findings on the effects of artificially reducing pitch error on expert violinists.

Chapter 9 continues with an in-the-wild study, using the augmented violin to provide an aural or visual guide with extra pitch feedback for helping pitch correction during lessons with beginners. During lessons students were asked to either use a pitch corrected aural guide provided through a single headphone on one ear, to look at a visual graphic highlighting pitch error, or a mixture of both or neither, in order to compare both the performance impact and the personal preferences towards the different feedback methods. The study also included a follow-up section testing the same artificial heard pitch error reduction as with experts but with beginners and in a non-laboratory environment. The chapter presents study design, execution, and results with user reaction to both additional pitch feedback mechanisms and use of artificial pitch error reduction.

Chapter 10 closes with a review of work completed and findings in Chapters 4-9. It also discusses future use and investigations using the augmented violin, especially the meaning of combined results from Chapter 8 and Chapter 9 and what they say about the viability of applying complexity management to violin learning.

Chapter 2

Learning and Practice

This thesis is motivated by the desire to improve student learning and practice motivation, so it is important to discuss theoretical aspects of violin and music tuition. Any tools for learning must be designed to consider how and what a student learns, and they must draw from established pedagogy. While there are many schools of violin study, there are also some key universal elements of learning shared by many instruments including digital music instruments (DMIs) which are relevant for particular consideration in the course of this thesis.

We begin with some basics of violin performance in order to introduce the reader to violin terminology and performance tasks before moving on to some of the more universal ways a student learns instrumental performance. We focus on the importance of cognitive load, and the means to reduce the cognitive demands of performance through *chunking*, breaking tasks into simpler sub-tasks, and internalization through practice.

Motivation, as discussed in Section 2.2.3, is essential for practice and therefore essential for learning. Without motivation, students do not practice, progress slows and instruments languish in their case. We review pedagogical guidelines for how to help keep students motivated before introducing the idea of *complexity management*; optimizing instrumental complexity so that the player's skill level can produce satisfactory musical results. Considering the initial complexity of the violin, we propose an augmented violin as a tool for experiments into complexity management.

As the learning studies conducted as part of this thesis focus primarily on pitch execution, we include discussion on the execution and learning of correct pitch. Lastly, we introduce potential ideas for relaxing pitch complexity by removing error from pitch performance in preparation for its application to the augmented violin (Chapter 7).

2.1 A Quick Introduction to Playing Violin

Although a full discussion of how to play the violin is beyond the scope of this thesis and would itself be subject to considerable debate, we include some basics in order to ensure readers are familiar with terminology and some fundamental aspects of playing.

2.1.1 The Violin and Bow

A traditional violin is depicted in Figure 2.1. It consists of a resonating *body* held under the player's chin with an attached *fingerboard* that hangs over the body and extends a similar length beyond it. The smooth rounded underside of the fingerboard beyond the body is called the *neck*. At the end of the fingerboard, farthest from the player, is the *scroll* and wooden *pegs*. The scroll is essentially decorative, but the pegs, housed in the *peg box* are used to adjust the pitch of each string. The *nut*, a small grooved piece of wood placed between the peg box and the fingerboard, is used to raise the strings above the fingerboard so that they can resonate freely. The grooves keep strings spaced apart. The other end of the strings are raised and spaced by the *bridge*. The bridge rests on the body of the violin and passes the vibration of the strings to resonating body. Strings are secured to the *tailpiece*. Tuning is most commonly accomplished by rotating the pegs, though many beginners have *fine tuners* where the string is secured to a metal tuner which is screwed to the tailpiece. Beginners typically use fine tuners on all strings though for advanced players fine tuners are only used with the E-String. The classical violin has 4 strings, G3, D4, A4, E5, each a fifth, apart with lower strings thicker than higher strings.

The vast majority of players use a chin rest and some sort of shoulder rest. The chin rest is clamped to the violin body and assists the chin in holding the violin in place. The shoulder rest, which can range in form from a cut-up sponge to an elegant carved wooden appendage, is placed under the violin to provide additional height to the violin so that it can be comfortably held through pressure between the player's chin and shoulder.

Figure 2.1: A violin (left) and bow (right) with labeled parts.

The bow, depicted in Figure 2.1 consists of a flexible wooden *stick* with horse *hair* anchored at either end. The bow is held at the *frog* with the far end called the *tip*. A standard full-sized bow is 750mm long with the hair 650mm long. The weight is typically $60 \pm 2g$ [176].

2.1.2 Basic Technique

The basics of the violin, as can be found in many books on learning violin [52, 91, 92, 85], start with upright body posture holding the violin between the left shoulder and the chin, rotating the head to the left. In order to deftly accomplish performance tasks and endure hours of practice, posture must be balanced and relaxed. In classical training, the violin is not supported by the left hand. Because the chin gently presses down on the chin rest, vibrations from the violin body can pass through the chin rest to the performer's jaw, giving the performer direct physical feedback.

The left hand is neutrally at rest towards the nut with the thumb one side of the fingerboard. Fingers are placed down on the string to change pitch, with higher pitch accomplished by placing a finger closer to the player's body. *Shifting* is the movement of the entire left hand out of neutral position towards the body to reach higher pitches. *Vibrato* is the achieved by rolling the finger so that the contact point oscillates around the primary pitch.

Figure 2.2: A sample bow hold. Keeping a relaxed bow hold with all fingers bent including the little finger can be a major challenge.

The bow is held in the right hand using a specialized bow hold depicted in Figure 2.2. Bow holds can vary significantly. Similarly, while modern classical technique requires holding the bow at the frog, some traditions hold it farther up the stick.

Sound is produced on the violin by drawing the bow across the strings. Tone quality and volume are primarily a function of bow speed, bow downforce and bow-bridge distance [2]. One of the most basic tasks for a violinist is to be able to "bow straight" meaning drawing the bow so that the hair is parallel to the bridge, there is no bow skew, and the bow-bridge distance remains constant. A second major task is to balance bow downforce and speed; too much downforce at a low speed will sound crunchy, whereas too little pressure can result in unfocused sound as the bow slides across, rather than grabbing the string. A violinist is expected to be able to use the full length of the bow hair, from frog to tip.

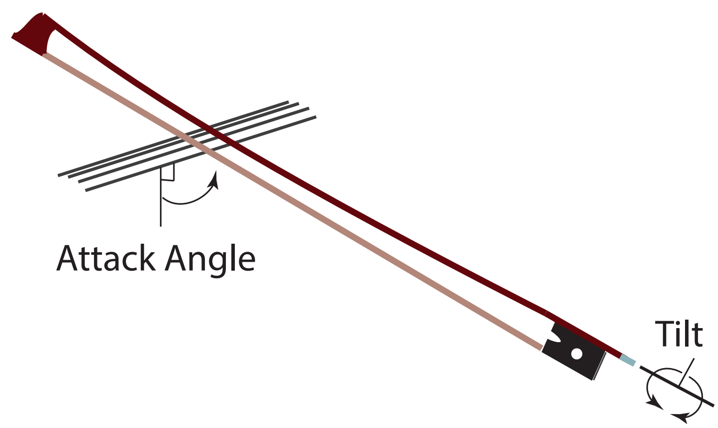


Figure 2.3: Bow 'attack angle' is the bow's vertical angle relative to the violin strings and determines which string is played. Bow tilt is the rotation around the bow's longitudinal axis.

Illustrated in Figure 2.3, two other bow parameters a beginning violinist must learn are bow *attack angle* and bow *tilt*. Bow attack angle is used to determine which string is playing. Though beginners will rarely play more than one string at a time intentionally, it can be very easy to catch a second string unintentionally. Bow tilt is the axial rotation of the bow as shown in Figure 2.3. Tilt and downforce determine the amount of hair contacting the string. Advanced players will almost always tilt the bow somewhat, using it as additional means to affect volume and sound quality, though for the beginning player, the task is more to maintain a constant low tilt.

The specific details of how a person reliably learns to accomplish left-hand fingering techniques and right hand control of the bow are not dealt with further here though they can be found in many books [52, 91, 92, 85] and through study with an experienced teacher. As renowned violin teacher Ivan Galamian wrote,

> I would like to point to the one-sided overemphasis on the purely physical and me-

> chanical aspects of violin technique, the ignoring of the fact that what is paramount in importance is not the physical movements as such but the mental control over them. The key to facility and accuracy and, ultimately, to complete mastery of violin technique is to be found in the relationship of mind to muscles, that is, in the ability to make the sequence of mental command and physical response as quick and as precise as possible. [52]

Following Galamian's advice, the rest of this chapter will focus more on learning theory, particularly as it applies to violin, rather than specific violin technique. However the reader should now be familiar with terms used throughout this thesis, as well as the basic technique tasks that users must practice and master to succeed at playing violin.

2.2 The Instrumental Learning Process

In depth research on exactly how a person learns an instrument is quite limited, probably due in large part to the long time periods, easily 10 years [41], involved for someone to transition from a novice to an expert. However instrumental teachers will often spend years teaching a student and guiding them towards proficiency. Though pedagogcial writings are often informed by personal opinion and learning theories instead of scientific study, considering the scarcity of existing research, it would be overly limiting not to consider the opinions of established respected teachers and their extended accumulated experience.

Similarly, though this thesis focuses on violin, much of the psychology, mental difficulty, and general physiological challenges of instrument learning are not violin specific, but shared across instruments. Hence we include references to not only respected string teachers such as Shinichi Suzuki, Edward Kreitman, Susan Kempter, but also wind teacher Daniel Kohut and guitar teacher Joseph O'Connor. Suzuki founded the Suzuki Method and philosophy of teaching which encompasses the music learning environment [165]. Kreitman writes about how to teach the fundamentals of musicality, posture, and practice to beginners [91, 92], while Kempter, who studied motor learning along with teaching, focuses on teaching physical mechanics and posture in violin practice [85]. Wind instructor Kohut [88], and guitar instructor O'Connor [124] both write about effective teaching, performance, basic music cognition, learning psychology and include existing research in their writing where it supports their arguments.

As Kohut [88] says, “No two people are exactly the same physiologically or psychologically. Thus no two people function or learn in exactly the same way. Consequently, there is no single process or mode of function that is ideal for all of us to use beyond the general principles inherent in the natural learning process.” Learning is an individual task and must be tailored to an individual’s problem areas and style of learning. However, there are common principles that are universally shared.

Learning an instrument is undertaken as a process with the end goal of musical understanding, capability or virtuosity. Musicianship is good for the brain. In a neuroscience review of the effects of music learning, Nina Kraus states, “Musicians show an advantage in processing pitch, timing and timbre of music compared with non-musicians. Music training also involves a high working-memory load, grooming of selective attention skills and implicit learning of the acoustic and syntactic rules that bind music together.” [90]

Musicianship is a longterm process because there are so many tasks to learn and coordinate that it would not be possible to learn them all at once. Musicianship requires not only individual cognitive tasks, such as being able to identify a note on a score, or placement of an individual finger on a string or key, but doing both tasks quickly and fluidly while also controlling the remaining fingers on both hands and listening for audio cues from other performers. Because of the demands on immediate recall, musicianship must be learned internally so that few individual tasks demand constant attention.

2.2.1 Cognitive Load and Internalization

Learning must be a multi-stage process. We are limited in how many things we can consciously handle at one time due to basic issues of conscious learning memory. Research by George Miller in the 1950s found that human short-term memory can deal with roughly “seven plus or minus two” discrete bits of information at any given time [118]. Subsequent research found that to convert those discrete bits into recallable long-term learning each discrete bit must be practised and repeated [85]. We will now discuss the concept of discrete bits, tasks, or "chunks," in pedagogy and how a chunk is learned.

Chunking

In a review of the cognitive neuroscience of music, Zatorre et al. [190] define *chunking* as "the re-organization or re-grouping of movement sequences into smaller sub-sequences during performance," and states, "chunking is thought to facilitate the smooth performance of complex movements and to improve motor memory." The implication of limited short-term memory and focus is that any large task achieved through the integration of many small tasks, must first start with internalizing the small tasks or *chunks*. The size of a learning chunk will depend on the sub-tasks required to achieve it and may be a combination of previously learned chunks or completely new ones.

Taking the hypothetical example of music reading, a foundational chunk is when the student must be able to identify that a dot (note) on the stave means a certain action. Playing the note requires the student to know how to execute the fingering action, though it does not necessary assume the student understands the note name or pitch equivalent of the note. Another chunk for the student is to learn how the drawing of the dot impacts timing and when to move from a first dot to a second dot. For a single dot, we've already identified two tasks to learn: placement and timing.

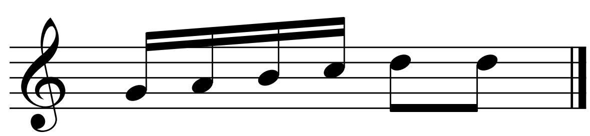


Figure 2.4: Simple chunkable pattern for beginners.

Having learned the basic meanings of the individual dots to a reasonable level, a student can then build upon that foundational understanding. For instance, taking an example from Kreitman [92], rather than thinking of Figure 2.4 as four similar length short notes followed by two longer notes, the student can link the two to a single larger block: e.g. the mneumonic"bu-sy, bu-sy, stop, stop". The student can also recognize that the pitch pattern is four ascending notes, and with an understanding of key, either conscious or unconscious, can chunk the pitch pattern as four up then a repeated note. Despite the fact the student is playing six notes, we still have only four chunked tasks: the rhythmic unit, the pitch sequence, and the evaluation and correction tasks.

The usefulness of planned chunking is the intentional focus on manageable learning to avoid cognitive overload. O'Connor's experience is, "Overload in either the number or complexity of the chunks results in difficulty and frustration. However naturally a student takes to the instrument at the beginning, it is worth resisting the temptation to add new topics." [124, p.52] Appropriate sizing and pacing of chunks is key to three things: successful learning, a sense of achievement, and continuing motivation. Managing task learning is not just for beginners but experts too. One of the teachers involved in our second study in Chapter 9 commented, "One of my big complaints with teaching in the [higher] academies is teachers either give you all this technique to work on while studying complex pieces and it is too much, or they put you on scales for a year so you can focus on technique and you are miserable."

Task Learning

Having recognized there are a limited number of tasks requiring short-term memory that we can focus on at one time, we must ask how we internalise a single task so we can use it as a building block of a larger chunk. There are four stages of perceptual-motor learning [76][1]:

1. unconscious incompetence
2. conscious incompetence (or conceptual understanding)
3. conscious competence (or voluntary action)
4. unconscious competence

In unconscious incompetence, a student is unaware they are not correctly performing a task. For instance, a completely new student may start by holding the violin up at the neck with

[1] Though a widely accepted idea, there are a number of variations in the specific naming of these stages, for instance though all three authors Johnson, O'Connor, and Kohut [76, 124, 88] discuss the same core concepts, all three use different names.

the left hand. For classical technique, supporting with the left hand is incorrect because it is detrimental to longterm performance goals, but for a beginner it seems like an obvious way to support the violin [85]. Once the teacher has pointed out how to correctly support the violin through the chin and shoulder, the student transitions from unconscious incompetence to conscious incompetence. They know they are holding the violin incorrectly, but have not yet learned to hold it beyond understanding the verbal instruction. As the student practices and is reminded by the teacher to hold the violin correctly, the student moves to conscious competence; the student can correctly accomplish the given task when directing focus at it. After continued repeated practice, the task reaches unconscious competence. The student holds the violin correctly without thinking about it.

The main goal of teaching is to aid a student's learning by aiding or introducing conceptual understanding and motivating the student to complete the next two stages for internalizing a task. Unconscious incompetence requires teacher intervention but the second two stages, conscious incompetence and conscious competence are where learning happens both with and without the teacher. Unconscious competence is the stage at which the learning task can be safely used as a task block that can be built upon for larger chunks. Having identified the need to separate learning targets into understandable tasks and the mental process through which a task must progress for proficiency, we must now consider how a student progresses a task from conscious incompetence to unconscious competence: practice.

2.2.2 Practice

It takes 3,400 hours to reach ABRSM grade 8 [160] and the average music student has practiced for 10,000 hours by the time they graduate from conservatoire [41]. Learning rates are indeed variable but largely due to effort. Shinichi Suzuki [165] says,

> Those who fail to practice sufficiently fail to acquire ability. Only the effort that is actually expended will bear results. There is no short-cut. If a five-minute-a-day person wants to accomplish what a three-hour-a-day person does [in three months], it will take him nine years.

A study of instrumental practice by John Sloboda supports Suzuki's observations. It found high standard deviation in achievement rates after a given number of years study, but also found that the raw number of hours required for instrumental progression was largely the same across both high-achieving and low-achieving individuals. Students who practiced more progressed faster.

However, there is a caveat to the idea that practice leads directly to progress; the practice must be correct. The classic phrase “practice makes perfect” is a misnomer: more accurately, “practice makes habit”. Some of the biggest outliers in the Sloboda’s study [160] were actually high-achieving students that took up to four times as many hours to complete a grade than average. The suggested cause of this is highly repetitive but ineffective practice blamed on inattention and repetition of error that must then be unlearned.

As an example of ineffective practice, violin teacher Kreitman tells a story where he asks a student who played terribly in a lesson whether she was listening as she played. She responded no, at which point he pointed out everyone else had had to listen to her poor playing and if they had to listen, wasn’t it only fair she also had to listen? So next time she listened while playing and played far better. If she persisted in practicing without listening, she would have failed to identify and correct error and would be bound to continuously repeat and habitualize bad performance [92].

The effect of practice is even visible studying brain development. Neuroimaging has revealed structural changes in the brain as a result of music practice. In particular, musicians show greater volume of the auditory cortex [190] and “that expert string players had a larger cortical representation of the digits of the left hand.”

The Basic Practice Loop

Kempter states good practice is effectively a closed-loop system [85]. The student has a target, thinks about that target, how to physically enact the target, performs the action, evaluates the results, refines the physical action necessary to achieve the target and tries again. Kempter [85] proffers a generic closed feedback loop for effective practice repetition leading to learning shown in Figure 2.5. Kempter focuses on the muscular actions involved in playing that must be internalized.

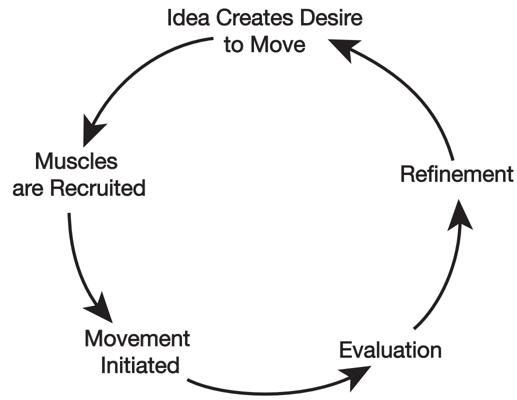


Figure 2.5: Kempter's closed-loop system for motor learning.

Key to the practice loop is not just the attention required to make assessment and refine the action, but also sufficient knowledge to assess correctness. As learning is accomplished through repeated executions of these loops whether the student is correcting or not, Kempter argues real-time assessment and refinement are more productive than occasional or retroactive correction. Real-time correction is more likely to prevent the development of bad habits [85, p.76].

A contrasting opinion comes from the academic Graham Percival. Using the analogy of a baby learning to walk, he suggests real-time feedback becomes a crutch: that we learn to respond to the feedback rather than the actual error, and that without the feedback, we are back to the beginning, a baby unable to walk. Although there is no available research to support either hypothesis, Kempter's real-world experience as a teacher and understanding of motor learning suggests potential value in her arguments, while Percival's arguments highlights potential risk.

Returning to cognitive load and practice loops, practicing often includes multiple loops, for instance, one assessing bow performance, one assessing posture, and one assessing pitch. In the common case of a student working on multiple tasks, it is not possible to always focus on every practice task. O'Connor suggests that to balance this, as the rate of forgetting is highest in early stages, it is necessary to practice new tasks on a daily basis with continued focused repetitions [124, p.58].

Designing Practice Aides

It is not always possible to practice fully correctly, even with motivation and attention, In Kempter's learning loop from Figure 2.5, the student must be able to assess correctness. In a lesson, the student has a teacher present to point out or intercede when the student plays incorrectly. During practice, the student must self-evaluate. Especially in the early stages of learning during conscious incompetence, it can be hard not just to properly correct an error, but even identify the error to begin with. Additionally, with an instrument like the violin, where even the most basic performance draws on multiple techniques, it would be impossible or impractical to learn every necessary technique separately without the student needing to use some aspect of another yet-to-be-learned technique.

For example, a soft flexible bow hold with a correctly angled loose wrist is crucial for the smooth and nimble bow changes demanded in advanced performance. However for the beginner, focus on right hand tasks is likely to be on the basics of holding the bow correctly, keeping fingers anchored in the correct places and the bow moving straight. While practicing these early tasks, students often let the bow hold become very stiff and fixed, acquiring a bad habit. Flexible bow hold is a new fundamental they haven't yet been taught to focus on. Kempter [85, p.76], focusing on the muscular aspects of early learning points out:

> If musical outcomes are the only stimulus which drives the muscles, they may or may not adopt beneficial movement patterns, which leave students at risk, Furthermore, dividing the limited resources of students' working memory between new, sometimes uncomfortable, movements and musical outcomes (rhythm, meter, phrasing) may not be possible at this time.

With the lack of a teacher, but still the need for assistance, well-defined technological practice tools may be able to fill the gap. While Section 3.3.1 looks at existing technological additions to practice, we briefly consider some main design criteria for technology assisted practice aides.

Rose Johnson, who studied the use of real-time vibrotactile and visual peripheral in real-world violin teaching scenarios, proposes a number of design guidelines for real-time learning aides in [76, p.127]. Excluding guidelines applicable only to her particular feedback methods, these are:

1. Feedback Focus: choose the feedback focus with the learning context in mind.
2. Modality:
   (a) Choose modalities to take into account individual differences in how people respond to different modalities.
   (b) Choose a multimodal system rather than communicating too much information in one modality.
3. Conspicuousness: fit the level of conspicuousness to the context and the individual.
4. Complexity: choose a single feedback focus.
5. Public Visibility: take into account social context when designing publicly visible feedback.
6. Metaphor: use embodied metaphors to describe feedback.
7. Negative vs. Positive: both negative and positive feedback is effective.
8. Forcefulness: feedback can be designed to be forceful or optional.
9. Context: take into account the demands and stresses of the setting.

Together, one of the main themes is to keep learning aides focused on a specific learning context and target. Johnson points out that many laboratory studies of real-time feedback in learning are performed where the participants are allowed exclusive focus on the task at hand. However, when people learn a skill in real life, there are often several competing tasks needing attention, meaning real-world learning tools must take the broader context into account.

A learning aid too broadly focused can be overwhelming and distracting. Focus and simplicity as a theme is in line with the use of chunking to help manage cognitive load. If targeting multiple areas for feedback, consider using different modalities for the different tasks. If possible use a different modality for feedback than those heavily used in performance. Modality choice is a way to keep feedback mentally separable and if more than one task is involved, it is better to split the cognitive load across many sensory skills than one.

Additionally when designing practice aides it is necessary to consider the best way to deliver feedback in order to catch attention without becoming a distraction. Johnson found many study participants reported feedback was better limited to specific error cases. For instance, when designing feedback, set an acceptable range as okay, only providing feedback when those boundaries were breached. As Johnson says:

> Violin playing should be an activity which places a high working memory load on the player. Since working memory load can disrupt a person's ability to focus his or her attention where they intend to, feedback will need to take the form of exogenous cues which can attract attention when necessary. Moreover, since participants may have difficulties mediating their attention between the feedback and other elements of playing there is a danger that the feedback may become a distraction if it continues to present information to the player when it is no longer necessary for improving their playing.

Lastly, as pedagog Juntunen and researcher Huanhuan highlight [79, 68], it is important to ensure any learning aide, especially in the augmented context, helps technique applicable to the instrument. They point out the popular computer games *Guitar Hero*[2] and *Rock Band*[3], where the user's task is to 'play' the target score using instrument controllers mimicking the guitar or other band instrument. Though ostensibly games based on musical instrument play, the musical scores and the instruments are so simplified that they become more training in motor reflex and memorization. It is widely accepted that the skills required to succeed at Guitar Hero do not transfer well to a real guitar.

2.2.3 Motivation

O'Connor summarizes the average guitar student's attitude to practice,"It was repetitive, it isolated and concentrated on the difficult sections. It aimed at improvement and was not much fun. It was something you had to do in order to play well. It was almost a chore that had to be done to make progress towards a promised land where they would be able to 'play the guitar' "[124, p.114]. Yet as we have seen in Section 2.2.2 practice is essential

[2] https://www.guitarhero.com/
[3] https://www.rockband4.com/

to instrument proficiency and musicianship. So how is practice motivation encouraged and sustained?

"Getting students to practice faithfully has always been a major concern of music teachers" [88, p.89]. As Kohut states, there are two types of motivation, which itself can be defined in this context as the reasons for acting or behaving productively towards practice and instrumental progress: internal and external [88]. Internal motivation is the inherent motivation and drive of the learner. Shinichi Suzuki places much emphasis on encouraging internal motivation by inculcating an appreciation of music and using parents and fellow students to make practice a socially desirable thing to do. Suzuki aims to set a supportive culture of music tuition that a student will then embrace [165].

As technologists, internal motivation is not something we can directly influence, but, following Suzuki's example, we can harness external motivators to influence students' internal motivation and having motivated students, can provide effective technologies for productive practice. Literature on instrument learning tends to be written from the perspective of how best can an adult, be it parent or teacher, motivate a child to practice and develop good practice habits [165, 91, 92, 85].

External motivation comes through outside incentives that are not self-initiated. Kohut [88] gives ten guides for external motivation. He writes these from the perspective of a teacher but in this thesis we are looking at technology to assist in learning and affecting motivation. Hence, removing teacher specific suggestions we are left with eight recommended ways to help maintain practice motivation:

1. Praise students when they are successful and don't criticise all the time.
2. Achievement must be measurable and both short-term and long-term progress should be verifiable in concrete ways. If possible, provide positive feedback when deserved.
3. Design a learning system with specific performance goals to work toward: provide them with a sense of purpose. Avoid trying to teach too much at once. This only leads to confusion and frustration in most cases, and even to despair in others.
4. The best motivator should be the music itself. Provide interesting and challenging music for them to play, for practice as well as for performance purposes.

5. Ensembles help. Being able to produce music with others can be a vital musical learning experience as well as an enjoyable social encounter.
6. Performances are helpful for providing a reason to practice and improve.
7. Provide balance between study and practice of fundamentals with real music.
8. Log practice time. While practice logs can be of questionable use as they create a connection that more practice is better, a practice log is helpful to establish the habit of regular daily practice.

We add one more, originating from Shinichi Suzuki, who was well known for teaching very young students: present practice and improvement as a game.

Combined, these mean that potential technology should be designed to provide positive or neutral feedback, and not highlight error so strongly as to be demotivating. Having various stages of achievement built into the technology could provide positive motivation to practice, the same way computer games have clear level goals. Technology must not add additional overly complex overhead to an already cognitively demanding process, rather technology that simplifies a practice task, either through highlighting error, or assisting correct user action, is more helpful. Technologies should be applicable to and assist with practice of real music and real performance, though targeting specific exercises can be helpful too. Further, if technology can facilitate or mimic ensemble play, it offers positive interaction and may make the tool more exciting to use. Lastly, having useful means to track work and progress over the longer term is generally beneficial. If designed well, technological teaching assistance can make learning and practice more enjoyable, and effective, increasing students' practice and instrument motivation.

2.3 Complexity Management

In this section we introduce the idea of complexity management: intentionally altering the inherent difficulty of an instrument in order to assist practice motivation. Complexity management is primarily relevant in designing new digital instruments but new technologies for augmenting instruments also allow the application of complexity management ideas to traditional instrument learning. We look at traditional means for discussing learning and performance progress on various instruments, including DMIs, before moving on to the idea of complexity management itself. Kohut points out,

> ...most people, regardless of training, can differentiate between extremely poor and good performance. Even young children are aware when their performance doesn't sound good, when the tone is poor and the intonation painful. This can be, and often is, a negative factor in terms of motivation for continued practice after the first two or three months of study. Good students in particular become frustrated if the tonal results they achieve are poor. But once tone and intonation improve sufficiently, practice comes easier, because it is more fun when one sounds better. [88, p.30]

Kohut suggests the period between starting an instrument, when a student produces unsatisfactory and poor results, and the point when enjoyable sound can be made and "practice comes easier" as lasting only the first two or three months of study. But Kohut writes as a wind player and teacher. Violin pedagogy suggests the frustrating hump before being able to reasonably play basic tunes is closer to two years [89, 85, 91].

Having looked at some of the basic aspects of music learning theory, it is valuable to acknowledge that not all instruments have an equal learning process. Some instruments have fewer input parameters to master or can be easily and cleanly chunked while others, such as the violin, can not. Further, accepting that practice is essential for expertise, it is necessary to consider the effect of a particular instrument's learning process on practice motivation. With this in mind, perhaps it is not surprising that few teenagers and adults take up study of the violin.

2.3.1 Learning Efficiency

Musical instruments must strike the right balance between challenge, frustration and boredom: instruments that are too simple tend not to provide rich experiences, and instruments that are too complex alienate the user before their richness can be extracted from them [97]. David Wessel refers to this balance as design for "low entry fee with no ceiling on virtuosity"

[177].

A traditional frame for comparing instrument learning is through a learning curve. Intuitively, the learning curve should map the expected relationship between cumulative applied effort and demonstrated musical skill within the context of that instrument. In his influential theoretical writings on digital lutherie, Sergi Jordà [78] reframes instrumental progress in terms of efficiency, a combination of the freedom the performer has to shape music and the potential complexity of the music they can produce, versus the complexity of input control. He calls this *music instrument efficiency* described by the pseudo-mathematical equation:

$$\mathit{Music\ Instrument\ Efficiency} = \frac{\mathit{Music\ Output\ Complexity} \times \mathit{Perf.\ Freedom}}{\mathit{Control\ Input\ Complexity}} \tag{2.1}$$

Though output complexity vs. input complexity loosely represent a more conventional idea of efficiency, music instrument efficiency includes performer freedom as a variable. Elaborating why it is included, Jordà gives the example of a CD player. Hitting play triggers a complex musical output with a simple low complexity input, yet no one would consider a CD player an interesting instrument as there is no effective performance freedom. An instrument must allow a musician to sound good or bad without which there is no opportunity to improve and no motivation for study. Music output complexity reflects the variation and range of outcomes available to the player, and control input complexity reflects the degrees of freedom in input and gesture that produces those outcomes.

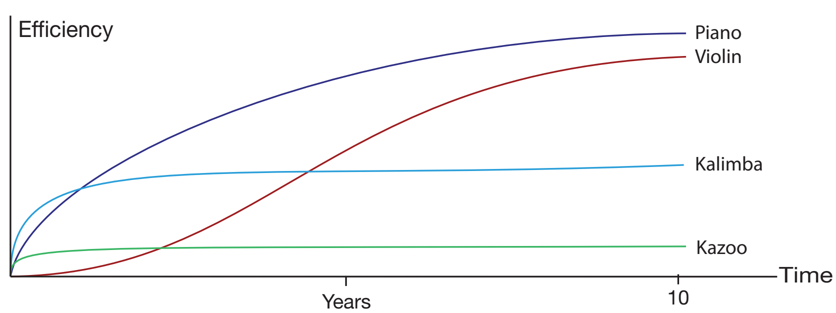


Figure 2.6: Sergei Jordà's learning curve based on music instrument efficiency [77].

As depicted in Figure 2.6, a major difference between the more traditional instrumental learning curve and a music instrument efficiency based learning curve is that commonly the learning curve is drawn within the context of potential learning on the given instrument, i.e. virtuosity is a common maximum across instruments, whereas music instrument efficiency lets us compare expressive virtuosic potential between instruments. Some instruments have inherently more potential for musical efficiency.

According to Jordà, efficiency changes with experience and practice. In Figure 2.6, we see the violin has very low starting efficiency as it has high input complexity, but inability to control inputs means low performance freedom and simple output. However, in the hands of an expert, the violin has music instrument efficiency exhibiting an exceedingly rich output palette and high performance freedom. A kalimba is much more efficient with low experience thanks in part to reasonably simple starting controls, but the efficiency curve flattens due to limited output complexity and the instrument never reaches high efficiency. The guitar also can be played with simple starting inputs. Virtuosity at the guitar results in arguably less performance freedom than the violin (which can technically be similarly strummed and has more intonation options), but more output complexity thanks to its polyphonic nature giving it a theoretical music instrument efficiency curve somewhere between the violin and the kalimba[4].

2.3.2 Designing for Learning

Though our examples of learning curves and music instrument efficiency all use traditional instruments, Jordà's arguments on music instrument efficiency are directed at DMIs. His ideas were developed in response to the challenge of designing new instruments. All the factors in music instrument efficiency, input and output complexity along with performance freedom are already fixed in traditional instruments. New technology enables us to revisit and intervene in traditional instrument learning but designing new instruments or augmenting technology does not necessarily let us throw out established pedagogy.

One of the oft-cited rules of instrument design is that large amounts of input control do not necessarily increase expressivity [39]. Referring to 3D motion capture, Bevilacqua [7]

[4]We are not arguing that one instrument is necessarily harder to play virtuosically than another, musicians will always push their respective instruments, only that some instruments are initially harder. Nor are we arguing that one instrument is in anyway better or more beautiful than another.

states, “A general problem with using a motion capture system as a controller for musical processes is that at 900 events/second ($\sim$ 30 points at 30 Hz), there are simply far too many to translate directly into a coherent musical process.” 30 control points for each moment in time does indeed seem like a lot, but that is only if we expect a learner to master all 30 points simultaneously and treat each point as a separate system. However, if we use traditional learning methods of chunking, repetition and internalization, with practice, 30 points may turn out both naturally connected and easily manageable.

Still, just as in traditional learning, if we fail to consider practice motivation, we can not expect a student to successfully learn control of those 30 points. As we have seen in Section 2.2.3, with motivation and instruction, a student will happily carry on to the level of expert, however without some basic musical rewards, a student is likely to abandon the task. Instrument complexity is very often talked about in DMI communities [77, 71], but strategies for addressing complexity stop at instrument build, rather than learning and repertoire. It is in learning where DMIs can have an inherent advantage over traditional instruments. In traditional instrument study, practice and learning are where strategies for dealing with complexity start. With augmented instruments and DMIs, we have the opportunity to consider complexity at all stages.

2.3.3 Reward for Practice

A major flaw in Jordà’s idea of defining instrument quality and learning curves through musical instrument efficiency is that it fails to incorporate the learner and the process of learning. Figure 2.6 implies learning through time, but discussion of instrument efficiency is limited to the the mechanics of the instrument rather than human motivation and enjoyment, the roll repertoire plays in influencing instrument gratification and progress, desirability of the end output, or the interaction between input capabilities (i.e. ability to chunk effectively). Without these additional factors, the idea of musical instrument efficiency is useful only in enabling a best guess of an instrument’s hypothetical end potential, but is ill equipped to illuminate how a player gets from beginner to virtuosi. As discussed in Section 2.2, recognizing the necessity of practice and the accompanying practice motivation for the success of any instrumental study, rather than considering an instrument’s basic learning curve or musical efficiency, we find it more useful to explicitly consider the amount of practice and work required to achieve a basic musical success.

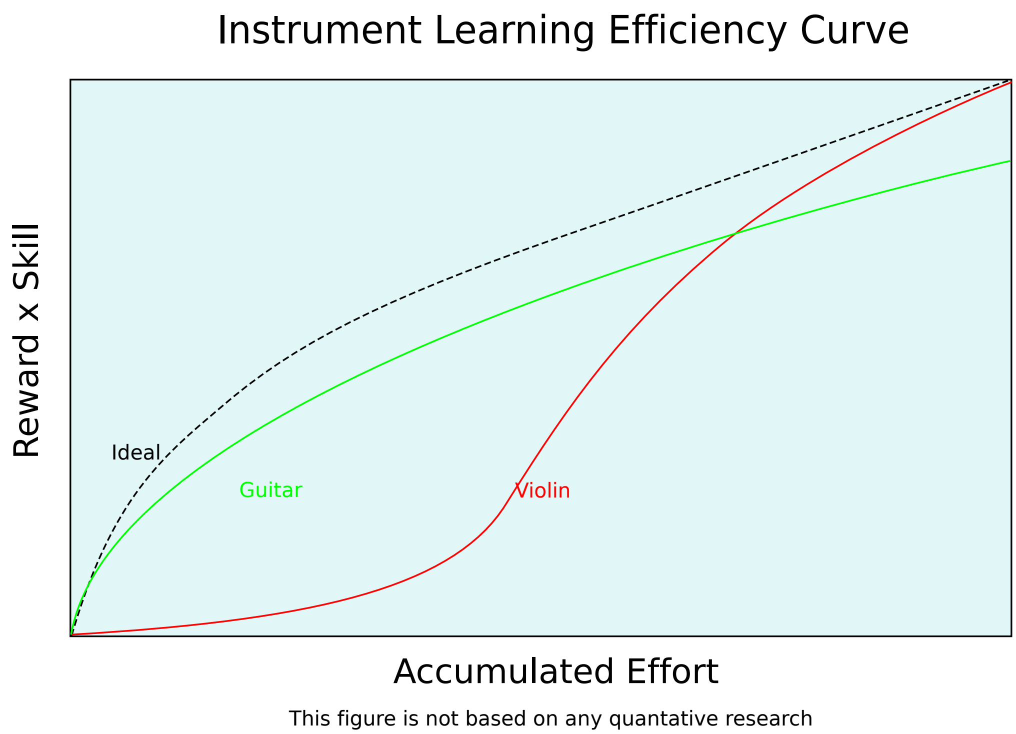


Figure 2.7: Sample expected learning and reward curves for the violin, guitar, and an ideal.

Figure 2.7 is illustrative of possible relationships between effort, and skill and reward for the violin, guitar, and an ideal instrument. Similar to Jordà's reconsideration of the learning curve and his idea of music instrument efficiency [77], rather than the traditional learning curve of effort versus skill, we use a learning efficiency curve based on instrumental skill, which is related to music output control and control input complexity, but also add musical reward. The pair more closely link to motivation than skill alone. Although improved skill is a motivation in itself, we believe that in the vast majority of cases, enjoyable musical outcome while developing skills is necessary to keep the performer interested. Music output reward takes into account whether the results of learned output controls are sufficient to be musically satisfying. Applying Csiksentmihalyi's (1996) concept of 'game flow' to instrument learning, if an instrument is too easily mastered, even with interesting musical reward, the performer will lack challenge and become bored. If too hard, musical rewards are low and progress is too frustrating. The ideal instrument is somewhere in between with instrumental skill and musical reward continuously increasing in relation to learning effort applied. Giving a similar pseudo-mathematical equation to Jordà's:

$$\textit{Learning Efficiency} = \frac{\textit{Music Output Complexity} \times \textit{Perf. Freedom} \times \textit{Musical Output Reward}}{\textit{Control Input Complexity}} \tag{2.2}$$

2.3.4 Making an Instrument Inherently Easier (or Harder)

With its lower maximum music instrument efficiency, why is the theoretical learning efficiency curve for guitar, as depicted in Figure 2.7, typically better than violin? The difference is in the quantity, difficulty, and separability of the techniques required to produce rewarding music. Unlimited pitch control and the continuous excitation from the bow mean the violinist will eventually have more expressive potential at his/her disposal, but for the beginner, the guitarist can quickly strum to produce a reasonable well-pitched sound. The guitarist can slowly add in complexity– e.g. chord difficulty, how many strings are played, different strumming techniques – while the violinist can not even simplify intonation. The ability to limit complexity while still achieving satisfactory musical results is potentially a significant reason for the guitar's better potential learning efficiency curve and uptake in later life.

Given the importance of exposed complexity to the learning experience, can we create an effective learning path through the management of complexity? If we can optimize the learning process, so an instrument is optimally complex for the user skill level, can we maximize practice reward thus improving motivation for long-term instrumental practice and instrument success?

To fully clarify what we mean by the idea of complexity management, here is a theoretical example: the fretted bass is far more popular than the fretless bass. This is possibly because the frets reduce the intonation options making correct pitch easier even though the fretless bass has a more expressive pitch range. Pitch-wise, the fretted bass is low complexity while the fretless is high complexity. If we had a variably-fretted base, we could slowly increase the number of frets as the performer demonstrates improved finger accuracy. The player continuously retains their ability to play predominantly in tune, but they slowly gain increased pitch-complexity and by the end, they have developed the skills to play the more pitch complex fretless bass.

Although pacing of difficulty and complexity is considered a major design aspect in computer gaming, applying the concept through instrument designs whose behaviour changes over time has not been well received amongst DMI designers. In his list of key DMI design rules, Perry Cook's second rule is "Smart instruments are often not smart." Cook [30] states, "learning from [a musician's] play and modifying its behavior often does not make any sense at all, and can be frustrating, paralyzing, or offensive." However, performable systems have been created that do not rely on pre-determined mappings and may indeed be "smart". Dave Merrill created *Flexigesture*, a tool to explore how different users might arrive at and prefer different mappings for a single interface [116]. Rebecca Fiebrink extended this idea by building an interactive toolbox where performers can choose input and output modes and then use machine learning to create real-time mappings [42].

Complexity management focuses on longterm practice motivation which means any in depth study of complexity management would be similarly longterm. The long term nature, and need to keep participants committed for an extended duration necessitates careful design of an appropriate test instrument, test simplifications, and preliminary studies to validate basic hypotheses such as 'beginners will enjoy this simplification.' Important initial questions in support of the primary question, can we improve learning efficiency through the management of instrument complexity, include how do we successfully alter complexity in a non-disruptive way, how fast do you add complexity, and would reduction of complexity habituate poor technique? As such, a full investigation into how learning motivation is affected by intentional management of complexity is beyond the scope of this thesis. However, we propose complexity management as a powerful idea for interface learning and begin the groundwork and preliminary studies necessary to better investigate the effect of varying instrument complexity to best improve user experience dependent on user skill in future.

Rather than work with a new DMI with no performance history, tradition of success (or failure), or repertoire, we lay the ground work for studying complexity management through the augmented violin. As an instrument with a very poor learning efficiency curve due to high levels of unavoidable complexity, the violin is a good candidate for potential learning process improvement. Violin technique is well established so we can judge outcomes in relation to normal methods and there is a substantial pool of people who want to learn violin. Provided skills developed playing our augmented violin transfer to general violin playing, the likelihood of attracting and sustaining a test group is much higher than with a novel interface. As part of this thesis we develop an augmented violin appropriate for experiments in complexity management and begin investigations into potentially effective means of reducing complexity through simplifications to pitch.

2.4 Learning Pitch

Intonation is a crucial learning task; wind pedagogue Kohut and string pedagogue Kreitman [88, 91] both put intonation third on their longer lists of teaching priorities. Orchestral stringed instruments are nearly unique in providing the player almost no physical aides for playing specific pitches. Winds and brass must adjust pitch through embouchure to be fully in tune, but the instruments still offer discrete means for hitting near a target note. Strings have no keys or frets, only a smooth fingerboard. The ability to play a target note specifically in tune is a challenge that can only be tackled through intensive practice and repetition. Kreitman puts emphasis on intonation as an aural-tactile relationship. Good intonation starts with good pitch sense and then development of the proprioception to execute the pitch.

2.4.1 Internal Pitch Sense

Before we can learn to place a note, we first must be able to hear it internally. Pedagogs suggest pitch understanding uses two two types of listening, external and internal. External listening is listening to what is actually coming through our ears: what is coming from outside. Internal listening is a mental image based on what music should sound like based on our aural memory. Kreitman suggests Edwin Gordon's term *audiation*, as an aural equivalent to visualization to describe internal listening[91, p.41].

> We use the auditory sense externally (we hear the note) and check that it is correct by comparing it with an internal auditory memory of what we expect to hear from our musical experience and sense of relative pitch..... this internal auditory sense takes time to build, unless a student already has perfect pitch. Students will blithely play notes out of key that set a musician's teeth on edge, because they lack auditory discrimination. [88]

Interviewed in an article about how to play in tune in *Strings Magazine*, violinist and teacher Michael Martin states [145], "Good intonation comes primarily from inside the player's head. If the player isn't hearing–the word we use is 'audiating'–good intonation in their mind, it's really not going to come out of the instrument." One of the major ways Suzuki, and Kreitman encourage development of the internal ear, the internal sense of pitch, is simply by listening to music played correctly. Listening to a study piece is expected to reinforce the mental version of both the music being heard and the standard intervals within them.

We tend to take the learning of pitch for granted. Most musicians can remember that they had to learn how to read music and associate symbols with instrument actions to produce auditory results, but assume the ability to differentiate tones is innate. This is very much not the case. In teaching young children intonation, Kreitman talks about how a teacher must guide the student through six stages of intonation [91]. The first two early tasks in teaching young children intonation are "can you differentiate between notes?" followed by "when they are different, can you describe whether the second pitch is higher or lower?" Though apparently trivial, young children must learn these tasks.

Studies also suggest pitch differentiation is a learned trait. Micheyl et al. [117] performed a study to investigate whether musicians were better at pitch discrimination than non-musicians. Using both pure sine tones and sine tones with added harmonics, musicians were more than six times better than non-musicians at differentiation of pitch (14.82 cents vs 2.25 cents) with an even greater discrepancy when listening only to tones with added harmonics. Additionally, pianists, who do not need to tune their instruments, were 33% worse than string and wind players who do need to actively tune (3.46 cents vs. 2.59 cents using pure sine wave). A last finding suggesting pitch is learned is that it took the average non-musician only 4-8 hrs of intentional training to be able to reach similar levels of pitch distinction as musicians. Non-musicians with training were unable to improve beyond musicians. The study did not look at whether non-musicians retained the learned pitch distinction beyond the main study.

Development of an internal ear has a clear impact on auditory cognition. In studies looking at the effects of music learning on brain development, Kraus reports,

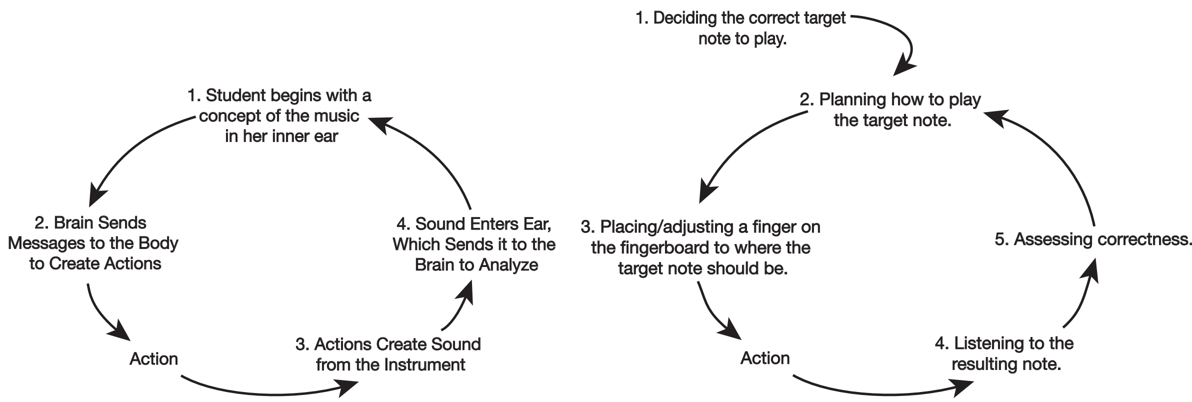
> Music training induces an enhancement of the processing of auditory signals, the characteristics of which depend on the nature of the training... on the practice strategies (for example, musicians who learn "by ear" show superior auditory encoding of musical sounds relative to those who rely on non-aural strategies) and on behavioural relevance (for example, the upper note, which often carries the melody in Western music and evokes a stronger neural response in musicians, or the emotion-bearing segment of a baby's cry...) Musicians are better at multi-sensory integration and audio stream-separation and better sound to meaning relationships. [90]

It is only once pitch is understood that a student can we move on to playing in tune.

2.4.2 The Intonation Process Loop

Playing in tune is a continuous process of improving pitch. It requires proprioception to physically target where the note is, which is then adjusted based on aural feedback. Aural feedback is not only used to refine correct location, but also to train correct placement in the first place. Kreitman [92] proposes a *listening loop* depicted in Figure 2.8(a) for tonalization. Looking back at Kempter's closed loop system for motor learning in Figure 2.5, we can see Kreitman's listening loop as much the same but with a more specific inclusion of the inner ear and listening. Here we further expand on both loops with a focus specifically on intonation. Performing correct intonation can be broken down into the six steps in Figure 2.8(b).

Each step requires different learned skills. The decision what note to play is commonly made by: knowing what the tune is and then hearing the next note, reading a score and knowing the dot on a particular score line is fingered a particular way, visually tracking the change from the previous note and matching the change physically, or knowing the note name. Additionally, once a piece is well learned, this and the next stage can become internalized so that motor



(a) Kreitman's Listening Loop

(b) Intonation Process Loop

Figure 2.8: Kreitman's listening loop for performing correctly with good tonalization [92] (a) and the pitch-focused Intonation Process Loop (b).

memory takes over. The first step from Figure 2.8(b), relating to note identification takes place outside the intonation process loop. Unless the player has made a mistake, it does not need to be repeated.

The second task is deciding how to play the planned note. Is it a third finger on the A-string or a second finger? Does it require a shift? Or are we correcting an earlier placement? Just as in step two and three in Kreitman's listening loop, the brain makes the decision and then instructs the muscles to carry out the actions needed to accomplish the target. Without frets or keys, the initial execution of where to put the finger is essentially based on vague guessing or learned motor memory developed through repetition. At the completion of step three, with the finger on the string, the pitch is played.

Next, the violinist listens to what they have played, step four of both Kreitman's listening loop and our intonation process, and then, in step five, the violinist must evaluate if it is correct. The decision whether it is correct is based on a mental comparison of the external pitch heard from the violin with the internal pitch imagined, or audiated, inside the player's head. The necessity to be able to hear correctness means that even when playing based on a score in step one, there must be a learned linkage between visual task and heard result.

Lastly, if the violinist has decided the performed pitch is inaccurate, they must correct it.

Again, they must plan how to play the target note, step two in Figure 2.8(b). Correction requires a decision based on prior pitch learning whether the performed pitch is too low or too high, by how much it is 'off', and which direction the finger must move to rectify the error.

Steps 2 through 5 are repeated until the correct target note is found and form the pitch feedback loop. After deciding the target, the finger is placed, pitch assessed, finger moved, and pitch assessed again and again. The cycle of adjusting the finger and assessing the result continues until either the target note is hit, the note duration is finished, or the performer has given up. Building a strong quick feedback loop is essential and fundamental for playing in tune.

There is a famous story about renowned virtuoso violinist Jascha Heifetz. When asked "How is it possible that you never play a note out of tune?" Heifetz responded, "I play many notes out of tune, I just fix them before anyone else hears them." [92] Teacher Michael Slechta [145] says, "When I hear [my students] hit a note and slide the finger into tune, I know what I'm doing is working because they're listening to themselves."

2.4.3 Practicalities of Tuning

Even for the experienced player, playing pitch perfectly is still difficult. Minor pitch alterations can affect how the violin resonates and, if playing in an ensemble, affects the harmonic blend. Additionally, without frets, similar to singers, it is possible for pitch to drift unintentionally; each note is largely correct in relation to the last, but not correct in relation to the open strings. This drift becomes painfully apparent once an open string is used.

One way to improve intonation, which also improves tone, is to use violin resonance and octaves to assist the internal ear. Kreitman refers to these notes as 'ringing tones' [91]. Simple examples include matching the first position third finger with the string below, or matching the first position fourth finger to the open string above.

Finally, there must be the recognition that intonation is not fixed, but dynamic. Though we stick to using equal temperament with a 440 Hz A in this thesis, for reasons both of simplicity and that it remains a good starting point for beginners, high-level performance is not so strict. Equal temperament is a compromise designed to make keyboard instruments capable of playing equally well in any key with 12 notes to the octave, a compromise not necessary with the violin. Writing about the history and evolution of intonation, Duffin states equal temperament is essentially the 'least worst' option [40]. A discussion of equal temperament in comparison to expressive intonation or meantone is useful to highlight that 'in tune' is not in truth absolute. There are other real-world pitch variations. For instance, period orchestras play to a 415Hz A while modern orchestras often use 442Hz. Add to that that sometimes other instruments a violinist might play with are just out of tune. The key for a violinist is to be able to listen and react. As Galamian says,

> Lastly, in this discussion of intonation, it is necessary to consider what type of intonation ought to be used: the tempered or the natural. This is not the place to go into the technicalities of the two systems. No violinist can play according to a mathematical formula; he can only follow the judgment of her own ear. Be this as it may, no one system of intonation will suffice alone. A performer has to constantly adjust her intonation to match her accompanying medium.
>
> The artist must be extremely sensitive and should have the ability to make instantaneous adjustments in her intonation. (The best and easiest way to make such adjustments is by means of vibrato.) An intonation adjustable to the means of the moment is the only safe answer to the big question of playing in tune. The most important part in all of this is assigned, obviously, to the ear, which has to catch immediately the slightest discrepancy between the pitch desired and the pitch produced and then demand an instant reaction from the fingers. [52]

2.5 Variable Pitch Correction

We are interested in whether we can make learning the violin more rewarding at early stages when the beginner is struggling to deal with the number of complex inputs. As pitch is one of the harder immediate tasks, we are interested in how a learner might respond to simplified pitch. An obvious simplification is to *snap* pitch so that it is only possible to play pitches in the chromatic scale just as the keys of the piano or the frets on a guitar. The learner still has to put their finger close enough to the correct pitch to select it, but we alter the output pitch to be in tune. But there is the question whether the student will find snapped pitch easier and also whether not being allowed to hear error limits learning.

Looking at aural feedback during performance, Bruno Repp [146] reported that studies removing aural feedback from piano performance did not find significant degradation in performance accuracy, but Repp's work with the pianists found loss of pitch feedback did have subtle impacts on expression. However the piano provides discrete pitch choices with substantial visual and haptic clues not available on the violin. It seems questionable to expect findings from removing pitch feedback on the piano to apply equally to stringed instruments.

A more recent study by Chen et al. looked at the effect of feedback when shifting on the cello [25]. Shifts were performed while using the bow, providing aural feedback, and without the bow, where there was no aural feedback. Chen found that though cellists were accurate when having aural feedback, cellists were only able to approximate the note without feedback.

If the output is always in tune, the intonation process loop is interrupted; the user is both unaware their input is out of tune and also how their input relates to the target pitch. We negate much of Kreitman's [91] second stage of learning intonation, "Can you describe whether the second pitch is higher or lower [than the first]?"

As a balance between always being in tune but unable to know actual pitch accuracy versus being able to determine inaccuracy but discouraged by the difficulty of correct intonation, we propose experimenting with changing the pitch input/output response so it can be designed to improve, but not fully correct pitch. We introduce the *pitch curve*, a function relating pitch input to pitch output, and call application of the different pitch curves *variable pitch snap*. *Pitch snap* refers to the use of a specific pitch curve. Further, we characterize each pitch curve with three parameters: *shape*, *strength*, and *speed*. Full implementation and algorithmic details of the pitch curve can be found in Section 7.2 but here we introduce the concepts behind variable pitch snap.

To better illustrate what we mean by pitch curve, we start with the normal pitch[5] relationship

[5]While pitch is sometimes equated with frequency, in the context of our discussion, we will use the term pitch to refer to the linearized version of frequency based on the steps in the chromatic equal-temperament scale quantified in cents, the linearized space between two chromatic steps.

for a violin. On a standard violin, there is a geometric relationship between the physical fingered position and the output pitch with continuous variation between two notes.

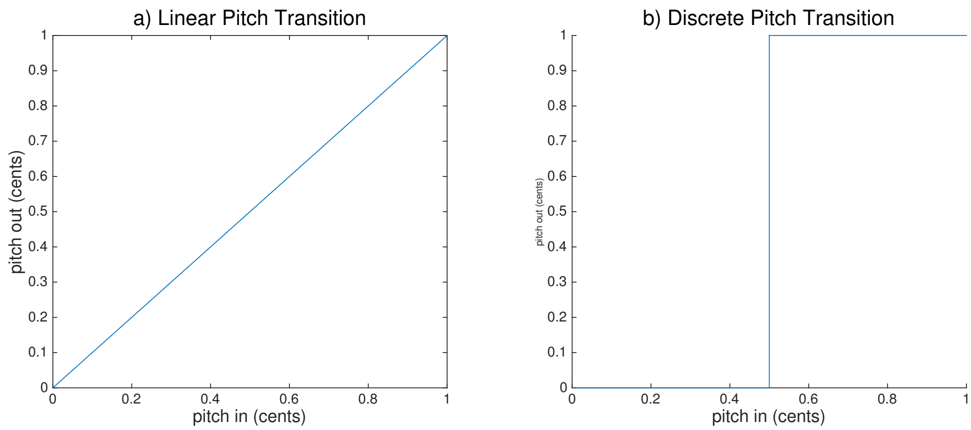


Figure 2.9: Two common pitch input to pitch output relationships. We are using a linearized version of pitch based on cents with 0 cents being an in tune note from the chromatic scale and 100 cents being the next in tune step in the chromatic scale, a semitone above. 50 cents effectively sounds halfway between two notes in the chromatic scale. The discrete change in (b) effectively rounds the input pitch to the nearest correctly tuned output so, for instance, a note performed 40 cents high is pulled to the perfect 0 cents pitch while with (a) the performed pitch in remains the performed pitch out, so an input 40 cents sharp remains 40 cents sharp.

The violin's natural in-out relationship is depicted in Figure 2.9 (a). We call it linear as output matches input and it can be modelled using a line. The synthetic full snap in (b) behaves similar to a piano, where the physical fingered input location only triggers distinct discrete note changes. With the augmented violin, we can arbitrarily chose to reshape the physical input to pitched note output using a new relationship of our choosing. For instance, in Figure 2.10, we have created new functions for physical input to pitched note that provide a response in between the violin's natural linear relationship (Figure 2.9a) and a piano's discrete sound steps (Figure 2.9b). In Figure 2.10 (a) has a disjoint transition while in (b) and (c), there continues to be a continuous link between input and pitched note. In all three the pitched output remains closer to the well-tuned note for physical inputs near the correct linear placement for the tuned note. This relationship between the actual performed pitch input and the augmented heard pitch output is the pitch curve.

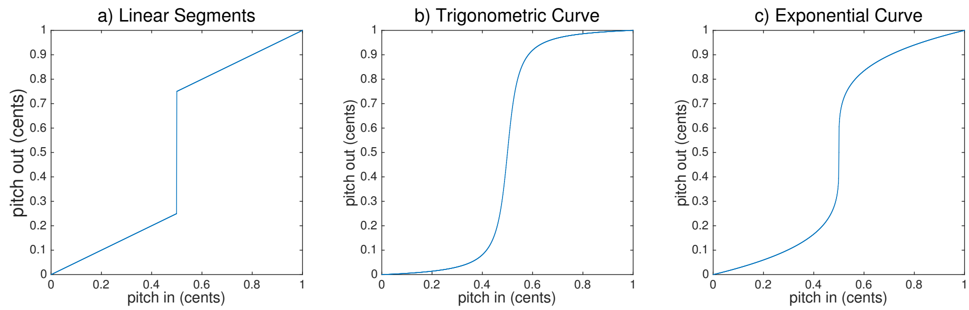


Figure 2.10: Three pitch input to output curves of different shapes. (a) demonstrates the linear segmented shape which is effectively disjoint when transitioning between snapping down vs snapping up. Shape (b) is a trigonometric arctan function which transitions more smoothly and (c) is an exponential curve that transitions more sharply than the arctan but is not as steep elsewhere.

2.5.1 Shape

In the context of a pitch curve, shape is defined as the character of the function determining pitch between two notes. Technically, we could use any arbitrary function with a unique solution; it could be linear segments, quadratic or sinusoidal segments or exponential. For illustration in Figure 2.10, using the linearized pitch of the performed input, the user can remap the output pitch according to either linear segments (a), a trigonometric function (b) or an exponential function (c). We are interested in learning whether the shape of pitch curve has any noticeable hearable affect on output. For instance, using linear segments, the transition between two notes can be disjoint. We would like to know the disjoint nature negatively impact sound and feel.

2.5.2 Strength

We defined strength as the degree to which the pitch is adjusted. In other words, strength lets us vary the level of pitch correctness. Sticking with the exponential curve as it varies smoothly and evenly, strength can be visualized by looking at the three varying pitch snap strengths shown in Figure 2.11. (a) is low strength which has been flattened to become linear or no snap, (c) is a strong snap so almost all audio output is near a scale note, and (b) is a medium strength snap where some deviation from the scale note is allowed but the output is pulled toward the in-tune scale note, reducing the deviation of the played pitch from the target note.

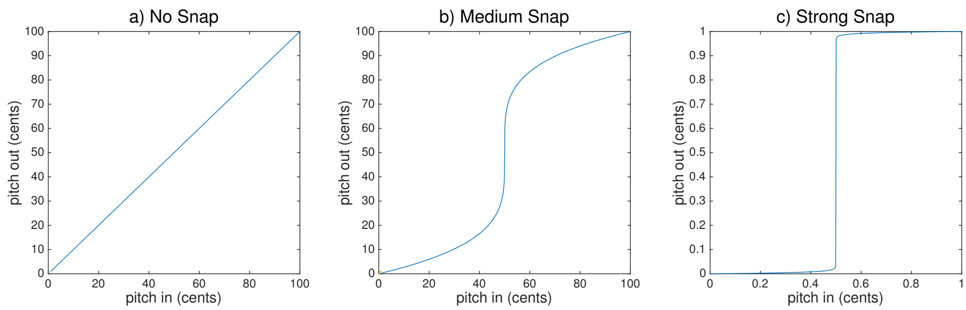


Figure 2.11: Different strengths of exponential pitch curve. (a) demonstrates low strength which is effectively no pitch change as output is the same as the input. For instance, a pitch input 40 cents above the nearest scalar semitone produces an output 40 cents high. (b) is with half-strength, where output pitch is pulled moderately to the nearest scalar semitone. Continuing with the input pitch 40 cents high, the output is now reduced to being only 17 cents high. In (c) the output pitch is pulled very strongly towards the nearest scalar semitone so now, for an input of 40 cents, the output is only 0.01 cents high.

2.5.3 Speed

Lastly, there is *speed*. Speed of snap is how fast the output is pulled from the original input pitch to the target snapped output pitch after a change in the chromatic scale note. Figure 2.12 shows three examples of the pitch switching from the a fixed input pitch, p_i, to a target

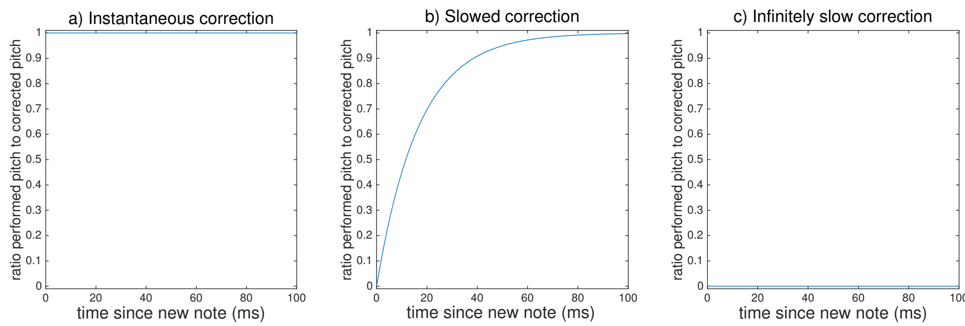


Figure 2.12: Different speeds of pitch correction. When a new note in the chromatic scale is detected, the ratio of the input pitch to the corrected output pitch will vary based on the time since the change. For the instantaneous change in (a), the output is immediately forced to the pitch corrected version while the for the infinite change in (c), the output is always the same as the input. For (b), it takes 50 ms to go from using the input pitch, to using corrected pitch at 95% strength.

output pitch, p_t. (a) depicts an instantaneous snap, (b) a slowly applied snap, and (c) an infinitely slow application of the pitch snap.

While the previous two parameters, shape and strength (and their accompanying figures) use linearized pitch input for the x axis, for speed, the x axis is now time. There are two reasons we have speed of pitch snap as an option. From an audio perspective, doing a snap immediately upon the note change often sounds unnatural as violin pitch transition is rarely instantaneous. As earlier discussed, violinists regularly slide into most notes, and as Heifetz pointed out, this slide can be barely perceptible yet it is a big part of how violinists tune the note. Thus, not only may immediate snap sound artificial, but feedback early in the note is particularly important to the performer's intonation process and removing it through pitch snap may be overly and unnecessarily detrimental. Lastly, as we will see in Chapter 5, the pitch estimation used to correct pitch is more likely to be incorrect at the very start of a note so that slightly delaying the snap may reduce snap the audibility of estimate error early on.

2.6 Summary

In this chapter, we covered a number of pedagogical and theoretical aspects of instrument and violin learning with a particular focus on intonation. We started with a short overview of violin terms and technique to ensure the reader has the necessary familiarity to understand discussion throughout this thesis. We then shifted to a less violin-centric discussion of instrument learning, discussing the long term nature of musical learning. We presented the idea of chunking for breaking down music learning into separable manageable blocks. Chunking is necessary because we have limited short-term memory and starting blocks must be internalized before we can build to learning more complex tasks. We also discussed the process with which a chunk, or learning task, progresses from unknown to unnoticed as the learner internalizes it. Studies confirm that repetitive practice is the main way the internalization process works, requiring self-reflection: knowledge, attention, listening, and most of all, motivation to actually do the necessary practice.

Seeing practice as the major contributor to instrumental success, we discussed the potential for technology to aide the practice process and help students assess their performance. We looked at Rose Johnson's work with learning technologies outside of the laboratory [76, 168] which reinforced that practice itself demands many mental resources meaning technological intervention must remain simple, designed with focus on a specific learning task and context, and must be thoughtfully designed to be forceful enough to cause a reaction, yet not overly distracting from regular play.

Our last aspect of instrumental learning was looking at how to encourage practice motivation, as without practice there is no progress. We found that there is both internal and external motivation, with external motivation having far more potential for manipulation. Clear measurable success on specific tasks, rewarding music, appropriate chunking of complex tasks, and monitoring of practice progress are some of the main ways to inspire student practice. Technological practice aides should be designed with awareness of these motivating factors.

Having argued not just the importance of practice and practice motivation, but the importance of achievable musical success and rewarding simplifications within practice motivation, we introduced the idea of complexity management: intentionally altering the inherent difficulty of an instrument in order to make it easier to achieve early success and maintain practice motivation. We look at the instrumental learning curve along with Sergei Jordà's music instrument efficiency curve [77] as descriptors for the music achievement over time but point out that neither curve includes musical motivation. We propose the idea of a learning efficiency curve which specifically includes musical reward and recognizes that high complexity instruments, like the violin, that can not be easily simplified, have very poor early learning efficiency as music making is overly difficult and frustrating.

The learning efficiency curve leads to complexity management as a way to use technology to pace instrumental difficulty with user skill level. We see complexity management as a potentially useful consideration in designing new digital instruments, not just enhancing traditional instrumental learning. Still, in this thesis we start to investigate complexity management using an augmented violin for the reason that the violin possesses poor early learning efficiency and already has a well established pedagogy and history of success.

As our experiments into use of the augmented violin as a practice aide and also our experiments into the implementation of complexity management with the violin are primarily centered on intonation, we return to look at how string players learn to pitch correctly. We focus on the necessity for a string player to develop an effective sense of internal pitch. Players must learn to hear when a note is correct and in tune and should be able to reproduce mentally or orally what music should sound like. It is only once a player has a mental knowledge of pitch that they can move on to actually performing pitch. We describe the intonation process loop, a derivation of Kreitman's listening loop [92], which focuses on the student hearing the note being played, comparing it with the mental target, physically correcting if wrong, and quickly continuously repeating till the note is correct.

Lastly, we propose the idea of correcting intonation as a means to simplify violin, but recognize that full correction breaks the natural intonation process loop. Listening is a key stage and if a student can not hear error, they can not correct it. We deal with this by suggesting partial pitch correction using a variable pitch snap, pulling a pitch toward the nearest selected scalar pitch. The behavior of a specific pitch snap is described by a pitch curve relating performed input pitch to heard output pitch. We propose three parameters, shape, strength, and speed as means to alter a pitch curve.

Chapter 3

Combining Violins and Technology

3.1 Integrating the Violin and Technology

Starting with Max Mathew's first forays into an electronic violin [109] in the 1970s, numerous digital instruments have been developed based on the violin family, including controllers inspired by the double bass and cello. Excellent in depth reviews of both design and use of bowed string controllers can be found in Poepel's and Overholt's writings [140, 126, 127].

Violin-based instruments utilizing technology are often split into three standard digital music instrument (DMI) categories as suggested by Marcelo Wanderley [172]: *instrument-like controllers*, *augmented instruments*, and *alternate controllers*. *Instrument-like controllers* are those that are intended to mimic an instrument, such as the EWI[1] or Yamaha SX-7[2], electronic wind instruments intended to play like the saxophone or clarinet. This category includes instruments that are inspired by existing traditional instruments but whose sound output is not necessarily related to the original instrument. *Augmented Instruments* are those adding sensors or actuators to an existing traditional instrument to either capture existing performance

[1] `http://www.akaipro.com/product/ewiusb`
[2] `http://windsynth.net/wx7_rev.html`

to manipulate the natural sound of the instrument or to control digital audio processing. (e.g [110, 8]). *Alternate controllers* are those that are largely novel and includes interfaces that do not fall into the other two categories (e.g. [107, 81]).

In this section, according to Wanderley's classifications, we highlight a number of different violin augmentations and violin-like controllers along with relevant design considerations for both. We also include a review of tracking systems used with the violin. These are considered separately from augmented violins since tracking systems do not involve designing any instrument alterations, just temporary placement of video markers or position sensors. Lastly, with our focus on learning, we look at the use of technology, including augmentations and violin tracking, in teaching and practice.

3.1.1 Augmented Violins

Augmented violins have the traditional acoustic or electric violin and bow as their primary starting point retaining standard playing techniques but adding new capabilities above and beyond the original. The augmented instrument may not look like the original, but still can be played like the original.

Augmented violins can be divided into two interrelated groups: instruments augmented to provide new performance techniques beyond the capabilities of the acoustic violin, and instruments altered for tracking of traditional forms of performance [126]. The difference in category is based on differences in how input relates to the original, rather than the output. For instance, Young's *Hyperbow* has been used both for examining bow techniques [187], and expanding performer gesture with novel output [184] but is classified as a tracking augmentation as it focuses on sensing traditional technique.

Augmentations Extending Technique

Augmented string instruments extending technique are playable using standard technique but also add a variety of new input modes useable to control electronic effects and sounds dramatically enhancing the range of expressive performance. Jon Rose motivated the first MIDI bow, working with STEIM in the 1980s to develop a bow with a pressure sensor controlled by the index finger and a sonar sensor to estimate bow position [126].

Interestingly, one of the earlier augmented instrument extending stringed technique was not based on the violin, but on the bass. Curtis Bahn's *SBass* [5] used an array of pickups placed around the body and bow of an electric upright bass along with a host of sensors mounted to the body, including a mouse-pad mounted under the fingerboard. Like the *Bowed Sensor Speaker Array* or BoSSA, a string-inspired instrument discussed in Section 3.1.3, output was through a set of spherical speaker arrays with mappings utilizing work from Perry Cook.

Continuing with lower strings, CNMAT built an augmented cello [50] in collaboration with cellist Frances-Marie Utti. The cello, based on an earlier electric cello design for Utti by Eric Jensen, allowed bowing above and below the string. The augmented cello design had a particular focus on tuning systems and included mechanical means for quickly altering string tuning. Multiple force sensitive resistors (FSRs) were used to provide continuous input sources with a switch array for selecting different settings installed under the bridge. A new bow interaction was created by adding a bowable rotary encoder below the bridge.

A less dramatically altered modern augmented cello is Dan Gibson's *Modified Cello* [55]. Like CNMAT's cello, Gibson included a specially designed rotary encoder along with a number of sliders placed on the body of an electric cello providing additional electronic controls.

Dan Overholt has developed two augmented violins for extending technique, the *Overtone Violin* [125] and its successor, the *Overtone Fiddle* [128]. The Overtone Violin is a redesign of the violin adding custom optical pickups for each string and inverting the normal design where tuners are at the scroll, instead placing tuners near the chin rest and adding a sensor block in place of the scroll. The sensor scroll includes a wide array of non-traditional inputs including buttons, sliders, sonar, accelerometers and more. Core to the Overtone Violin design was the idea that performance extensions are derived from adding new modes of interaction, rather than repurposing existing gesture.

Though the Overtone Fiddle looks less radical that the Overtone Violin, its design is still highly novel including the addition of transducers on the fiddle. It replaces the custom optical pickups with a commercial magnetic pick-up and is designed to be fully stand-alone. Based on performance experience with the Overtone Violin, the Fiddle replaces the large number of input sensors with a mounted iPhone. The advantage of the iPhone is it can be programmed based on the specific performance intentions, and can perform all necessary DSP tasks for generating output including driving on-instrument active transducers. A specialized wireless bow includes orientation sensors which can be compared with orientation sensors in the iPhone for additional free gestural input. An optional resonance box underneath the normal violin interface is used to amplify electronically generated audio.

Another instrument based on the cello is Halldór Úlfarsson's *Halldorophone 5*[3]. Úlfarsson's Halldorophones pick up the sound produced by the instrument and use it to electrically drive the same instrument, effectively feeding back on itself. The feedback loop alters the normal resonance of the instrument. For instance, on the Halldorophone 5, a plucked string can sustain infinitely. The Halldorophone 5 includes controls for setting the feedback to each string individually.

Before moving on, one of Perry Cook's design principles for DMIs [30] is that, "Some players have spare bandwidth, some do not." While this is normally directed at the physical and mental limitations of the player, the comparatively high number of cellos and basses in the category of augmentations extending technique suggests a possibly different spin on the principle; "Some instruments have spare space, some do not." Physically, the space on the violin and viola is limited for adding manipulable sensors without taking Overholt's Overtone Violin approach, essentially redesigning the violin.

Augmentations for Tracking

While so far we have looked at augmented instruments enabling completely new performance techniques and gestures, work in this thesis falls into the category of augmentations for tracking. A major reason to augment for tracking is for technical analysis of how people play though it can also be used for score following and/or novel performance applications. Traditional technique analysis focuses in two areas: left-hand finger placement (connected to pitch) and right-hand bow tracking. Bow tracking has been well explored [2, 188, 144, 131], though

[3] `http://www.halldorulfarsson.info/halldorophone5/`

it is by no means a fully solved problem.

Fingerboard Tracking

Fingerboard tracking is, in certain ways, an easier problem as the physical contact location between the finger and the instrument is directly measurable. Freed examined use of commercial position and force sensors [48, 51] for finger position tracking in a traditional context, while Grosshauser augmented a traditional violin with both capactive and pressure based position sensors [62, 60]. Ajay Kapur's *E-Sitar* [82] is a fretted instrument which uses a different fingerboard measurement method, linking the metallic frets with resistors and measuring string-fret contact by electrifying the strings. Although this is not a technique available to fretless instruments the E-Sitar is notable for using audio analysis to refine estimated pitch, as sitar pitch is determined by both the frets and the bending of the strings.

There are also a variety of guitar tracking techniques based on video tracking [21, 19] however typical techniques used, such as adding reference markers, or doing feature point detection are difficult to apply to the violin. Reference markers are easily occluded or awkward to mount due to the violin's small size and weight while, without frets or other clear fingerboard markings, the violin has a poor visual texture for effectively finding keypoints [35].

Bow Tracking

Effective bow gesture capture has been an ongoing target since Askenfelt's pioneering work in the 1980s. One of the challenges when characterizing bow action is the array of parameters that contribute to the bow-instrument interaction. As identified by Askenfelt and others [2, 151], there are seven main parameters available to a string player when bowing: 1) bow velocity, 2) bow pressure, 3) bow to bridge distance, 4) bow position defined as the transversal position along the bow, 5) bow tilt, 6) bow skew defined as the bow's angle to to the bridge, and 7) bow attack angle which primarily determines string. The first three are the most important for driving the acoustical response of the string while the remaining four allow the player to moew effectively control velocity, pressure, and position or add nuance to the tone.

Askenfelt's early work [2] sought to capture bow gesture by altering the bow, adding a thin resistance wire into the bow hair. By also attaching the bow hair to the bow through strain gauges, Askenfelt was able to capture position, pressure and velocity. Askenfelt later was able to capture bow-bridge distance by electrifying the strings to act as resistance wires [3].

Gershenfeld and Paradiso subsequently developed the *Hyper Cello* which tracked bow transverse position and bow-bridge distance through electric field sensing. This was accomplished by attaching an antenna behind a cello bridge which was used to drive a resistive strip along the bow [129]. The *Hyperbow* [184] is an evolution of the Hyper Cello and has received many upgrades over the years such as force detection, inertial capture, and wireless operations [186, 144]. Many of the best practices from the Hyperbow have been incorporated into the commercially available K-Bow[4]. A drawback of the approach of the Hyperbow and K-Bow is the requirement for a specialized custom bow and the difference in bow balance caused from the technological additions.

A particularly useful development in bow tracking was Demoucron's work [37] looking to physically model the interactions of the bow. In order to characterise bow performance and use, Demoucron developed an easily mountable force sensor based on a strain gauge mounted at the frog to measure bow hair deflection. When combined with accurate position, measurements from the strain gauge can be used to derive bow pressure effectively. Use of this method requires calibrating the system [64] by performing sample bow strokes allowing the system to account for differences in flex and performance of different bows as well as bow tension during use. Though some promise for capturing position through the addition of a second sensor at the tip was demonstrated [36], Demoucron's design is typically used with only the sensor at the frog and a separate high accuracy tracking system. It has enabled accurate capture of force in many applications such as identification and characterization of different bow strokes and performance styles [155, 156, 154, 63, 24, 26].

[4] `http://www.keithmcmillen.com/k-bow/`

3.1.2 Non-Augmenting Violin Tracking Techniques

There are two main means of violin tracking that we are considering separate from augmented violins, video tracking and similar vision based motion-capture systems along with electromagnetic field (EMF) position tracking. While highly effective, they are applications of non-specialized tracking systems, can be used with ordinary violins and bows, and do not require the same level of design consideration as true augmentations. For instance, pure video tracking, such as used in the iDVT [100], and De Sorbier's augmented reality violin [35], do not even require the addition of any sensors or marking.

Video motion capture is now a relatively common method for violin tracking and is used in a multitude of projects at different locations [155, 122]. It is often paired with Demoucron's strain gauge [36] to enable bow force measurement. Markers are easily added and are lightweight. Placement can be challenging as violin performance requirements mean some optimal marker locations are not feasible and playing will often result in some markers becoming occluded. The major drawbacks of video motion capture however are that systems are expensive (in the order of $15k plus), immobile without significant calibration, and require significant expertise to use. These make motion capture highly useful in musicological and pedagogical studies, but not appropriate for broader use.

EMF tracking, primarily used at UPF [26, 64, 24, 63], was first seen in Peiper's work [135] but was really brought to maturation by Maestre [101]. Modern EMF tracking systems use small wired sensors that can be attached to various points of the violin and bow. Maestre used a Pohlemus Liberty[5] EMF tracking system which provides full motion and orientation data. By placing a single small sensor on the bow and a second on the violin, Maestre et al. were able to accurately determine bow transverse position, velocity and the bow to bridge distance. The approach also enabled the ability to track the string being played and estimate bow pressure. Though there have been attempts to derive force from physical models using just EMF tracking [108], Demoucron's strain gauge design paired with effective position tracking still remains the optimal method for deriving force estimates. EMF is high accuracy, relatively easy to use, and does not suffer from the same occlusion issues encountered in video motion capture, however drawbacks are that sensors must physically connect to a base station

[5]`http://www.polhemus.com/?page=Motion_Liberty`

through wires and multi-sensor systems are expensive (in the order of $10k).

3.1.3 Violin-Inspired Interfaces

Violin-inspired interfaces are typically not actual violins, but instruments which take inspiration from the form and gestures of the violin. One of the most publicly well known violin-inspired interfaces is Laurie Anderson's tape-bow violins [161]. Anderson replaced the bow hair with magnetic tape and the bridge with a magnetic tape head so that the recording on the bow is played back through a bowing motion.

Another instrument used repeatedly in performance is Dan Trueman's BoSSA [166]. Though the BoSSA is radically different from a violin, trading the violin's resonating body for a speaker array, the interactions are violin-inspired. Four pressure sensors replace the traditional fingerboard interaction, and actuation is accomplished by bowing mimicked using an augmented bow fitted with bi-axial accelerometers and pressure sensors, the *R-Bow*.

There is also Suguro Goto's *SuperPolm* midi violin [57]. The SuperPolm again trades strings for synthesis with touch-strip sensors replacing the fingerboard and the bow actings as a voltage divider with the wiper being the bridge.

Lastly, the *O-Bow* by Dylan Menzies [115] is an input device controlled by a normal violin bow. The O-Bow uses the sensor from an optical mouse to track movement of the bow for bowing input to control a synthesizer. The O-Bow also uses slight alterations in bow angle to control vibrato.

3.2 Design Guidelines for Augmented Violins

Typically, a DMI is composed of an input interface, an output sound engine, and mappings between the two, though with augmented instruments the sound engine may be the instrument's acoustic sound. DMIs can take many forms based on a variety of input, output, and mapping styles. Paradiso [130] and Bongers [14] extensively detail input possibilities for new musical interfaces. Paradiso categorizes instruments based on usage modalities somewhat similar, but not fully aligned, with traditional instruments; namely keyboards, percussion, batons, guitars, strings, wind instruments, voice, wearables and non-contact sensing.

Although the augmented violin may retain its original audio output, it is still considered a DMI. Here we discuss aspects of DMI design that apply to augmented violins. Some issues, such as practical build, are relevant to any violin-based DMI, while others, such as mapping, are relevant only for augmented violins designed to expand performance capabilities.

3.2.1 Practical Build and Implementation

A frequent joke within the DMI community is that new DMIs work right until demonstration time. One of our core design goals is a practical augmented violin for use in repeated performance and studies, meaning practical issues are imperative. Cook succinctly summarized some of the key lessons for practical effective instrument design in [30, 31]. Practical design considerations are concerned with reliability, survivability, repeatability, backward compatibility, and ease of start-up. For instance, using wireless technology for data transfer may turn out less reliable than wires and similarly, batteries may run out at an inopportune time. Unfortunately, technical issues in prototypes relating to reliability are difficult to eliminate, but following best practice and incorporating a significant test period can help mitigate problems.

Portability and ability for reproduction are further important technical challenges. An instrument requiring expensive or cumbersome hardware will be restricted in use. For instance, motion tracking, popular in studying violin performance, has traditionally been limited for practical performance as, not only do motion capture systems require extensive setup and calibration [83], but both motion capture and the main alternative, EMF tracking [101], are prohibitively expensive. Jensenius discusses the challenges of portability and reproduction in a violin-based performance in [73] using simplified visual tracking.

Another major low-level practical consideration is latency. Latency can be described as the responsiveness of the system, or the time from performer action, to system reaction. Responsiveness of the system impacts the player's ability to exert exact control. Low latency responses that feel instant are much easier to understand and learn than high latency responses. Wessel and Freed suggest a target latency under 10ms [49, 177] with more recent empirical studies by Jack supporting that 10ms target. Jack found that a 10ms latency did not degrade non-professional musicians' performance or experience, though 20ms did [72].

Possibly more important is jitter, or variation in latency [177]. While a 25ms latency may be noticeable to a player, if it is stable, there is evidence the player can learn the expected audio delay [96]. However, if the latency is variable, the audio delay is no longer predictable and it becomes impossible for the performer to effectively compensate for any unwanted delay. Jack found that adding $\pm 3ms$ delay to the otherwise acceptable 10ms delay resulted in a significantly more negative experience for musicians [72].

3.2.2 Learnable Interfaces

Although Chapter 2 focused on learning and musical efficiency from a pedagogical standpoint, the role of proprioceptive and aural feedback and repetition should also factor into instrument and augmentation design. The 'human-machine interaction loop' [14], depicted in Figure 3.1, represents an expanded interaction loop similar to Kempter's closed-loop system for motor learning discussed in Section 2.2.2 but, as it is oriented at general DMI design, inserts the technological system as the recipient of the action. Just as the brain does, the technological system must also sense the human action, process the event, and actuate an output in response.

Wanderley distinguishes that there are two levels of feedback for the user in human machine interactions: *primary feedback* originating from the tactile-kinaesthetic or visual interactions with the sensors and instrument in order to play it, and *secondary feedback* based on the sound produced by the instrument [172].

A designer must consider primary feedback in order to evaluate learning feasibility. For instance, increasing tactile feedback by adding extra haptic interactions such as notching a violin-like fingerboard can improve the ease of learning in the same way frets help players physically find correct pitch on the guitar. Conversely, as will be explained later, a common design mistake is to use movement in space to control a multitude of fine discrete events [15, 47, 132]. Proprioception is not as refined as our somatosensory sense of touch [172]. This

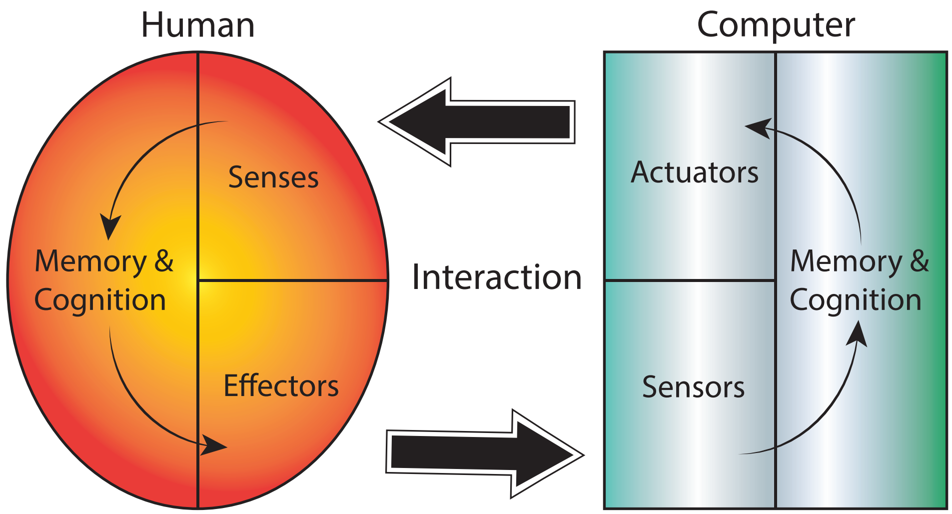


Figure 3.1: The Human-Machine Interaction Loop [14]

is not to say non-contact interactions are invalid, only that the inferred controls and outputs generated by non-contact interactions must take into account the (in)ability to reproduce input actions.

For secondary feedback, an impediment to learning is non-determinism such as latency and unpredictable generative responses. Additionally, inability to distinguish the effects of control inputs (i.e. insufficient volume or timbre space) can also impede secondary feedback effectiveness.

One of Cook's design principles is, "Smart instruments are often not smart" by which he means that instruments that learn from the player and alter how they perform can often feel unpredictable and invalidate earlier human learning making them "frustrating, paralyzing, or offensive." Complexity management includes the eventual potential target of a smart instrument pacing its own difficulty in apparent contradiction to this principle. However, Cook later revised the principle advising caution when designing smart instruments to ensure that the player is not overly confused or surprised by changes in instrument behavior [31].

3.2.3 Design for the Instrument People Play

Hochenbaum points out that most widespread music learning tools use a piano keyboard for interaction even though the keyboard has little relevance to many students' primary instrument [66]. Hochenbaum argues that augmented instruments are the way forward in instrumental learning and we should be designing for the instruments people want to play.

With augmented instruments for learning, the performer is expecting to use and learn normally applicable expertise, so augmentations should minimally interfere with most modes of play. Most importantly for any augmented violin use of the fingerboard must not be hindered and alterations to the bow, such as weight and balance, should not drastically alter normal performance.

3.2.4 Mappings

Much has been written about DMI mappings: level of control [174], the risk of repeated remapping resulting in a music controller that is more of an interface rather than an instrument [97, 177, 78], whether a mapping should always be predictable [103], focusing on constraints rather than affordances [103, 102, 189], evaluation techniques [14, 174, 139, 86], and effective design strategies [173, 69].

Though an in depth discussion of mapping is relevant to what we could do with our augmented violin, within this thesis we focus on augmenting an instrument for learning purposes. Our learning goals strongly dictate our mapping criteria, so we limit ourselves to a discussion of natural mappings.

In a study of instrument usage [103], Magnusson found that people disliked arbitrary mappings where mappings did not reflect "natural mappings between the exertion of bodily energy and the resulting sound." A musician has a concept of which gesture-sound actions make sense, and which do not. For instance, discrete outputs based on continuous non-differentiated input actions, like moving a hand freely through space, break the idea of 'natural mappings' as one small movement may cause nothing to happen while a very similar small movement might cause a lot to happen.

Similarly, though one-to-one linear mappings are the easiest to understand and implement, one-to-one linear mappings are rarely as effective in both performance and enjoyment as one-to-many or non-linear mappings which are closer to those naturally inherent in stringed instruments. Hunt and Wanderley performed experiments where users were asked to perform simple audio tasks using variations of one-to-one mappings and one-to-many mappings [70]. While users had an easier time determining the human-machine interaction when using a one-to-one mapping, they actually performed better when using the non-linear one-to-many mapping. Additionally, they tended to enjoy the non-linear response and find it more 'instrument like.

Rovan conducted similar mapping studies using a wind controller [149]. Considering that, like a violin, a traditional wind instrument has many cross-coupled input-output relationships, his experiments also included many-to-one mappings. He found the same preference for complex non-linear mappings in expert wind performers [71] and that these non-linear mappings were important in their ability to reproduce control similar to their traditional instruments.

3.3 Technology for Learning and Practice

Recent technological advancements have naturally made their way into violin study. Many of the augmentations and techniques described in Section 3.1 were developed with educational motivations. Here we look at the use of music technology in learning with a focus on interventions in the normal practice methodology for the violin.

There are a wealth of ways technology has been used to augment tuition. As discussed by Juntunen [79, 80], distance learning, online tutorials, remote masterclasses, video review, and audio play-along and are now available and often popular. These interactions are made possible by technology, but though they may alter how content is delivered, they do not necessarily alter the content. We are more interested in tools that provide new information or enable new approaches for how to learn: technology enabled practice tutors.

As emphasised in Section 2.2.2, practice is where the most time is spent learning the violin, but also the most isolated. Practice is where tools to help highlight error and provide feedback for how to improve have a particular opportunity to assist learning. As Juntunen states:

> Most students are on their own in home practice; their families don't know the instrument, cannot tune it or supervise the practice. The young students have no one at home to tell them whether they are playing correctly or not and it is also possible that all the time and effort ends up consolidating the wrong thing. As the saying goes, 'practice makes permanent', it will take additional time and efforts to bring a student back onto the right track again.

3.3.1 Non-Violin Learning Technologies

While it is not possible to provide a full review of all technologically-assisted instrument learning systems, we include a sample of some of the more significant projects along with examples across different instruments and feedback methods.

One on-line based learning project that goes well beyond simple video tutorials or video-conferenced lessons, is the *Practice and peRformance Analysis Inspiring Social Education* (PRAISE) project[6] [183]. PRAISE is a multi-instrument on-line music learning community which focuses on how to enable optimal feedback for learning. Designed based on observation and analysis of actual lessons, it also acts as a platform to explore new technologically-aided learning tools in pursuit of an effective learning environment. PRAISE provides a personal media repository for giving and receiving feedback that includes many of the more traditional web-based tools such as peer review, including a media timeline that is used to enable peer feedback on specific portions of a recording, but also features non-traditional feedback such as automatic audio and gesture analysis tools.

A novel aspect of PRAISE beyond many online learning communities is the inclusion of a software agent for technologically mediating learning tasks. The agent is designed to monitor student progress and suggest appropriate lesson plans or tutorials based on what the student needs to learn and their previously demonstrated skills [182]. The agent also proposes social connections for optimal assistance in a specific learning task.

One example of new technological interaction explored for use with PRAISE was the use of Kinect skeleton models to provide posture feedback [56]. We single this study out for its

[6] http://www.iiia.csic.es/praise/

inclusion of violists in the user study. Participants found that the virtual skeleton was only useful for basic posture feedback in both teaching and individual practice contexts.

Another learning tool that includes a significant social learning context is the *VirtualEuropeanMusicSchool* (VEMUS). VEMUS is the successor of IMUTUS [143, 158]. IMUTUS, designed for recorders, uses score following to enable automatic page turning when in practice mode, and has a performance mode that rates and analyses a performance based on various audio characteristics such as incorrect notes, fingering, or articulation. Errors are filtered and prioritised so only the three most severe errors are reported. Error filtering is intended to prevent a student from becoming discouraged when practicing with IMUTUS. The system includes advice for fixing an error and tools for easy import of new music scores.

VEMUS [46] features support for popular wind instruments such as clarinet, saxophone, and flute, and offers enhanced feedback such as visualization of audio features within a performance, and multiple means for annotating a performance. Notably, VEMUS adds support for use in social contexts, so while useable by a single student during practice, it can also be used by a teacher, either locally or remotely, during group or solo lessons.

While IMUTUS and VEMUS support an array of wind instruments, many technological learning aides focus on the piano, in large part due to the ease with which MIDI can be acurately derived from piano playing [136, 66]. One of the earlier successful piano-based instrument learning systems was Dannenberg's *Piano Tutor*. The Piano Tutor used score following for real-time page turning and accompanying. Multimedia feedback was provided after a performance using performance evaluation algorithms. Lastly, the Piano Tutor included a curriculum and a model to track student performance and move students through it. If a student demonstrated weakness in a certain area it would repeat lessons on that task, or drop to something more fundamental. An extensive user study found participants were not only positive about learning with the Piano Tutor, but also successfully retained the piano skills they developed through the study.

The systems discussed so far have predominately employed visual or peer feedback, with aural interactions limited to recording replays or audio examples. In contrast, Holland et al. [67] investigated using haptic and aural feedback to teach drummers four limbed drumming patterns. Holland attached simple vibro-tactile feedback devices to each wrist and ankle that buzzed for each percussive strike performed by that limb. Holland tested the effectiveness of vibro-tactile feedback in comparison and in conjunction with a repeated aural guide of the target rhythm. He found that drummers preferred aural over vibro-tactile feedback, but were still able to learn rhythms with only vibro-tactile training. The combination of both aural and vibro-tactile approaches yielded the best performance overall as participants found they encouraged different styles of learning.

Switching to guitar and visual tracking, Cakmakci designed a digital learning assistant for bass guitar fingering [21]. Using video tracking and score following, the system uses a virtual mirror and puts a dot over where the player is supposed to be pressing the string. The dot does not move to the next note until the player has placed their finger on the correct fret.

Lastly, Menzies designed an electronic bagpipe chanter for practice and pedagogical purposes [114]. The augmented chanter uses a breath sensor and optical sensor embedded in a 3D printed chanter to track breath and finger placements and can be used on its own with a computer, or can be used with a full set of pipes. The augmented chanter works as a practice enabler by allowing silent bagpipe practice with the full set of pipes but also goes much further, identifying fingerings and capturing ornamentation.

One of the defining characteristics of highland pipe playing is the specific ornamentation. Before a beginner can play even the simplest tunes, they have to learn a large array of complex ornamentation techniques that can take years to fully master. In order to assist beginners tackle these complex early learning tasks, Menzies built a system to identify and evaluate ornamentation. He including it in a game *Bagpipe Hero*, modelled after the popular Guitar Hero games[7], that rated players' playing accuracy. He tested the augmented chanter and ornamentation evaluator in both an extended one-on-one teaching setting and in individual practice. Users found it helpful for highlighting error and increasing awareness. In a study of independent practice using the augmented chanter, participants found it easy and insightful to use with all participants opting to use it regularly throughout the study.

[7] https://www.guitarhero.com/

3.3.2 Violin Specific Technology Enabled Practice Tools

One of the first significant attempts to build a violin practice and pedagogical tool was with the *iMaestro* system by Ng, Larkin et al [122, 121, 120]. The iMaestro system used a Vicon motion capture system for tracking violin performance. One of the main interactions, inspired by the use of a mirror in practice, was that performance actions could be viewed in a 3D setting with the player able to rotate views of tracking points in order to perceive performance actions from multiple new view points. The visualization included the ability to include bowing trajectories, along with bow position, bowing joint angles and more. Later versions included score following components [120].

Recognizing that while playing, visual attention is often required for performance tasks, the iMaestro system included sonification for real-time feedback [94]. Sonifications were tested both continuously and when specific events happened, for instance only if a player's bowing angle exceeded a pre-set threshold.

Despite the iMaestro's impressive scope, depth, and the inclusion of teachers in design stages, little is written about user tests. No formal use study was performed though it was presented to string teachers and students for limited testing. Limited tests found positive overall responses and viewed iMaestro as successfully demonstrating new information to users. Opinions of the sonifications were quite broad. Some teachers found them helpful, some distracting, though there was a general preference for discrete rather than continuous feedback. Students were generally positive, appreciating the ability to monitor their performance.

Similarly, Schoonderwaldt [157] has considered visualization of bow actions for pedagogical reasons. Amongst his many explorations using motion capture in conjunction with Demoucron's strain gauge [37], Schoonderwaldt presents two-dimensional visualizations tracking the bow's frog in relation to the violin along with a visualization of bow tilt. Combined with further visualizations of force and additional derivable information such as velocity, Schoonderwaldt proposes that such information would be useful for pedagogical and practice purposes. There is no discussion of user tests and Schoonderwaldt also describes accurate measurement of bowing gestures as "tedious" and acknowledges that the use of a Vicon system, also used in iMaestro, restricts the implementation to the lab.

Grosshauser proposed two ideas for performance feedback, performance feedback using vibro-tactile stimuli [58] and sonification [59]. In both cases, Grosshauser described using an accelerometer and gyroscope to sense aspects of player motion. In [58], he provided vibro-tactile feedback through a motor attached to the bow, while in [59] aural feedback is provided through two speakers attached either side of a music stand. For instance, if attempting a straight bow, the sensors could potentially detect incorrect lateral motion and apply vibration to the bow that would be sensed by the player, or alter the pan on a feedback tone based on the deviation level. No performance verification or user study was included in either work so it is unclear how effective the approaches were.

The *MusicJacket* by van der Linden et al [169] is based on using the jacket from a motion capture suit to track performer posture and then provide vibro-tactile feedback to encourage corrections. Vibro-tactile feedback was used as a feedback means as it is a relatively unused sensing modality and different placements can suggest different actions. Position data was used to inform feedback on bowing straight and correctly holding the instrument. Preliminary and follow-up studies suggested that though some users found it distracting, use of the MusicJacket distinctly improved straight bowing. In laboratory tests, the improvement only existed while feedback was on, with improvement stopping when feedback turned off, giving mixed evidence whether training with the feedback would improve unaided performance.

To follow up on laboratory results, the MusicJacket was taken out for an extended two-month trial in schools and used in actual violin lessons, a real-world natural use environment [168]. The MusicJacket proved successful, but in part because the implementation was flexible enough to allow feedback rules to be modified to focus on a student's particular problem area, otherwise it could became an unhelpful distraction. Also, especially working with young children, unconsciously incompetent, it was necessary to have the teacher teach both how to interpret and how to correctly respond to provided feedback.

Johnson continued work based on the ideas behind the MusicJacket focusing on tools to encourage use of the full bow length [74] and good posture [75]. The MusicJacket used an expensive tracking system unsuitable for practical purchase and distribution so these studies used simplified inexpensive tracking based on the use of gyroscopes to estimate bow location through change in lower arm angle and accelerometers to estimate incline in posture. In all studies, a mixture of vibro-tactile and visual feedback were used. Again, both tools were used in extensive user studies yielding mixed results. One of the main lessons is that the use and response to a given feedback method is often very personal to the user; while one user finds a flashing visual distracting, another finds the flashing necessary to draw attention. Feedback must also be appropriate to the working context. For instance, though one study demonstrated visual feedback useful to help participants use full bows, in a follow-up study it was barely noticed as other tasks vied for mental attention. A major difference was that in the first study, participants were asked to focus on bow use, whereas in the second, it was not as much of a focal point and participants instead focused on reading music.

While the majority of studies have demonstrated possible approaches for user interaction, it is difficult to draw useful lessons for alternative practice approaches due in large part to lack of significant evaluation. van der Linden and Johnson's and work with the MusicJacket and follow-on feedback tools [168, 76] offers not only excellent insight into potential practical design of effective practice tools, but also effective deployment of tools, and demonstrates that with limited study sizes, in real-world contexts, results using a given tool are likely to vary widely.

Methods for Pitch Feedback

Despite the fact that, so far, the feedback methods discussed have focused almost exclusively on the bow arm and posture, left hand tasks on the violin are also exceedingly challenging with minimal visual or physical references, and can benefit from feedback. We review four systems, three for violin, one for voice, designed with pedagogical motivations that are primarily directed at playing in tune using pitch tracking and score following.

The first is the *Interactive Digital Violin Tutoring* system (iDVT) by Lu, Zhang, Wang & Leow [100]. The iDVT uses an interesting audio-visual fusion for accurate transcription of violin performance. In order to mitigate problems with audio-only string transcription, the iDVT uses video tracking of the fingers and bow to supplement pitch and onset information. In [68] Huanhuan described use of the iDVT for players to self-assess performance. The iDVT transcribes a player's performance and then compares the transcription to a pre-recorded correct version with incorrect notes highlighted. Analyzed video is displayed as well. Though capturing correct notes, the iDVT does not relay more detailed information relating to whether notes are in tune. Additionally the iDVT is not a real-time system, receiving negative feedback in user studies for processing delays. Further, calibration for video analysis is tricky. System evaluation consisted of teacher feedback followed by trials with two students. Participants found the iDVT feedback "useful" and "considerate."

InTune [98] added richer pitch feedback to results similar to the iDVT. Lim and Raphael use audio analysis only to transcribe performance, solving problems with audio-only transcription by aligning the performance to a pre-existing score. InTune was designed for and tested using vocalists, but provides intonation feedback also relevant for a violinist and as it relies on audio input only, it is not necessarily instrument specific.

A non-real-time system, InTune displays the performed score along with a second piano roll style display of actual notes played including full pitch information. The performed semitone is highlighted and within it is the fine depiction of pitch as played. Lim and Raphael include a 20 person use test with highly experienced vocalists who found the system informative, especially with regards to vibrato. Interestingly, a common point of feedback was that the pitch feedback would probably be most useful for beginners when students are less familiar with correct pitch.

In [175], Wang proposes a violin-specific real-time tutoring tool for pitch, again based on audio transcription. Wang offers pitch feedback at a level of detail between those in the iDVT and InTune. Based on a pre-existing score, performed pitch is compared to the correct note and instant visuals are added once the note has been played depicting whether the note was correct, sharp or flat. Arrows are used to indicate whether too high or too low with the number of arrows depending on the degree of error. No user performance or evaluation was included.

It is worth adding that a significant drawback to all three of these pitch-based approaches, iDVT, InTune, and Wang's work, is that they rely on a score for ground truth. Though less of an issue when learning highly standardized repertoire, practicing a new piece requires finding or generating the reference score. This is often impractical and reduces both flexibility and practicality.

One pitch feedback tool taking a novel non-score based approach is de Sorbier [35]. Taking inspiration from attempts to design visual feedback for guitar playing such as [119] and [99] which both use visual analysis to create virtual fret patterns, de Sorbier used a Kinect to track the violin and fingerboard and then display a video of the player overlaid with virtual frets. In this way, the student can see whether their finger is placed in the correct location. Additionally, audio pitch analysis was added to inform arrows on the visualization directing which way the student should move their finger to correct to the nearest note. The violin offers limited space for attaching visual markers and has poor texture for visual differentiation so a Kinect was used to add additional physical information for tracking. Still, accuracy of the virtual frets was on average 7.22mm which, depending on location, is ± 28 cents error. Though accuracy tests were conducted, no user studies were performed.

3.4 Summary

In this chapter we discussed the two major goals when bringing technology to the violin: enhancing performance through new interactions, or tracking violin performance for pedagogical or classification purposes. We reviewed existing augmented violins and violin-inspired interfaces used for both enhancing and tracking purposes. We also discussed effective means for tracking violin performance using non-violin-specific tracking techniques. Work in this thesis is primarily pedagogically oriented with a focus on violin tracking.

Due to a combination of limited accuracy, expensive custom bows, and the alteration in feel due to adding sensors when tracking through augmentation, tracking for pedagogical purposes is predominantly done using non-violin-specific methods. These are typically based on either video motion capture [155] or electro-magnetic field tracking [101] in combination with a specialized spring mechanism for measuring pressure designed by Demoucron [37]. While highly effective, the major drawback is that these non-violin-specific methods are highly expensive and non-portable making them inappropriate for use outside the laboratory.

We subsequently discussed various requirements when designing augmentations, not just for tracking purposes, in order for an instrument to suceed. An instrument must overcome practical issues such as the need to be robust, and easy to use, as well as the need for an augmented instrument to perform in a manner that is predictable. To be predicable, an instrument must have stable low latency, and preferably utilize interactions naturally learned by the human brain and body.

We reviewed existing technological music learning systems. We included a sampling of non-violin-specific aids including networked tools that support stringed instrument studies along with other instruments, before discussing violin specific learning aids. Many aids, like iMaestro [120] and MusicJacket [169], utilize motion capture technologies to provide feedback on bowing and posture, which again, due to cost and complexity are inappropriate for broad distribution. Only Johnson has studied a variety of simple inexpensive tracking approaches for providing vibro-tactile and visual feedback to help address specific bow and posture tasks that have also been tested in real-world contexts [76].

In this thesis, we focus on the universally challenging task of playing with correct intonation. Existing work into intonation aids, for not only violinists, but also singers, are either reflective, providing visualization of when a performer is flat or sharp only after they have completed the performance, or are visualizations based on score following. Score following requires substantial setup and restricts the natural flow of practice by its linear nature. The only real-time pitch feedback tool we have encountered suffered from insufficient accuracy for effective use [35]. Additionally, all of these intonation training aids have so far been predominately focused on visual modes of feedback despite intonation being an aural task, as discussed in Chapter 2. The focus on visual interactions does not match with how violinists currently learn or approach intonation encouraging us to question whether or not it is the best interaction mode for a pitch aid.

Chapter 4

Bow Tracking Using Optical Sensors

This chapter incorporates significant material from 'Near-Field Optical Reflective Sensing for Bow Tracking' by Pardue and McPherson originally published in the proceedings for NIME 2013 [131] and 'A Low-Cost Real-Time Tracking System for Violin' by Pardue, Harte, and McPherson published in JNMR 44.4 [134].

4.1 Targets & Motivation

Tracking any performer's motions while playing a musical instrument is a challenging task, and this is especially true for violin-family bow technique. Bow tracking has been a frequent topic of research within performing and academic communities in areas including creative performance [148], studies of bow mechanics [3], and performance analysis [142, 152].

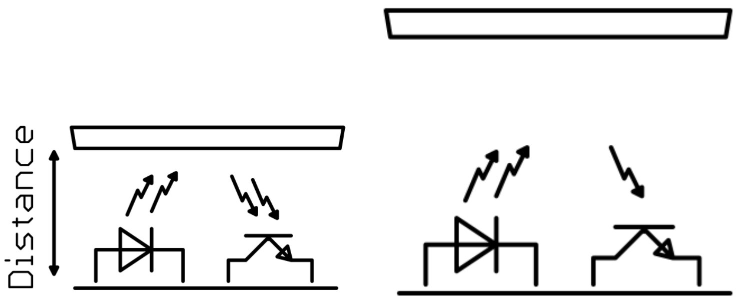
The ideal tracking system would let the performer play in any situation to the best of their ability, yet capture every musically relevant dimension at high spatial and temporal resolution. Optimal performance typically involves using the performer's own bow and instrument, and not impinging on performance freedom. A performer's relationship with their instrument is intensely personal and altering that relationship risks a performance. Similarly, physical additions or cabling that change the bow's normal feel disturb the musician. Unnatural restrictions on the performer's motions may both limit full body expression, but also mentally distract from the performance. In practice, every sensor technology will make compromises in one or more of these areas, altering tools or dictating the performance environment. Considering the potential value of the practical information available through bow gesture capture for bowed instrument practice, designing a system that limits physical interference, has high portability and low cost is significantly worthwhile.

As reviewed in more depth in Section 3.1.1, existing bow tracking options tend to be either expensive [101], non-portable [155], require a custom bow [110], have limited accuracy across one or many bow parameters [144], or a combination of disadvantages. Although the use of both video motion capture and EMF tracking with or without an added sensor for pressure sensing [36] has become a reliable approach for studies of bow technique they use highly expensive technology not commonly accessible outside the research environment.

In contrast, we focused on the most relevant demands for practical use of bow tracking in a generic learning context: cost, portability, immediacy of results, and intrusiveness. We were intentionally not looking to design the most accurate method of bow tracking nor to capture all possible bow parameters, but to design a real-time system that could be built affordably, mount on existing bows with limited paraphernalia, and achieve sufficient accuracy for use in a practical learning environment. In pursuit of this low-cost, low-intrusive, real-time bow tracker, we developed a new method of bow tracking using optical sensors.

4.2 Bow Tracking Though Bow-Hair Displacement

We combine optical sensors to track fine changes in the deflection of the bow hair and use that to estimate bow position and pressure. Optical reflective sensing involves shining a light source, typically an LED, at a reflective object and measuring the intensity of the reflection with a phototransistor or photodiode. They are small, lightweight, and require minimal circuitry, making them ideal for mounting in the limited space between the stick and the hair. As illustrated in Figure 4.1, they work by reflecting infrared (IR) light off a surface and releasing current in response to how much of the emitted light returns. Several manufacturers produce small inexpensive integrated packages containing both emitter and detector. More advanced digital optical reflectance sensors reduce noise from the ambient environment by modulating the IR signal at high frequencies. In contrast to video motion capture techniques, typical distances in near-field optical sensing range from 1-20mm, with micrometer-level spatial resolution.



(a) Close Object (b) Far Object

Figure 4.1: Basic principles of near-field optical reflective sensors. IR light bounces off a reflecting surface, and hits a photo-receptor. The amount of reflected light determines the electrical response. As demonstrated by the differences between a) and b), if the reflecting surface is farther away, the IR light will have more opportunity to disperse rather than hit the photo-receptor, decreasing the electrical response.

We apply near-field optical sensors mounted on the bow to enable high-speed tracking of bow position and pressure through measurement of the distance between the bow hair and the stick. When played, the bow hair is pressed towards the stick by the string forming the *displacement triangle* as shown in Figure 4.2. The location of the triangle's apex along the stick gives the bow position while the depth of the apex gives pressure.

One implication of the displacement triangle is that, for a given bow tension, every pressure and position combination gives a unique triangle and thus a unique measurement of that triangle. Even though the hair-stick distance at one point along the bow may be the same for different combinations of pressure and position, using the hair connection at the tip and frog as two known anchor points, the relationship of the hair-stick distances between two additional non-linear points will be a unique solution. This can be proved through a simple experiment demonstrated in Figure 4.3. We start by modelling the tip and frog as the two

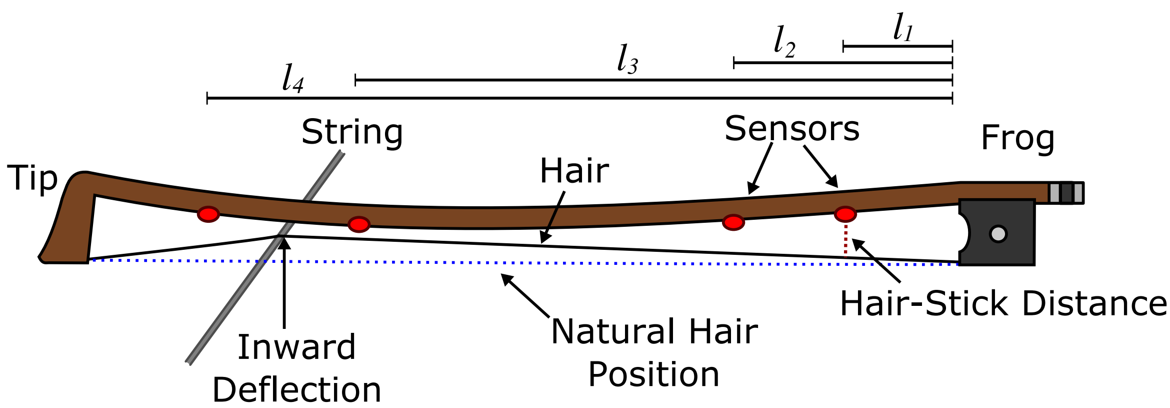
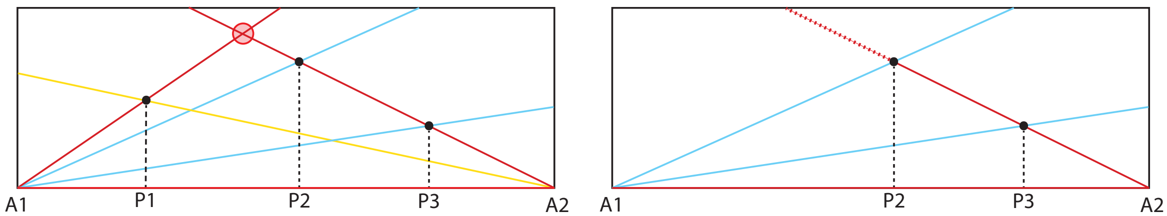


Figure 4.2: Basic mechanics of bow and hair deformation when the bow is pressed against the string. The string pushes the hair towards the stick. With the two ends of the hair fixed, the resulting hair forms two sides of what we term the 'displacement triangle'.

anchor points and initially add three measurement points(Figure 4.3 (a)). If a line is drawn from each anchor through each point, a triangle will be formed from the line between the anchors and the two lines that between them, run through all points and both anchors. The intersection of those two lines is the triangle's apex. Note that the intersection point would not change if we had only one of the measurements P2 or P3 as they are co-linear. However, if only given P2 and P3 as in Figure 4.3 (b), we can not confidently determine the left side of the triangle only knowing the triangle apex is somewhere along the dashed red line. Multiple optical sensors placed along the bow let us measure the hair-stick distance and through comparison of those distances, estimate which bow pressure and position creates that particular displacement triangle.

Bow tension is critical for pressure estimates since it will take more pressure to displace the hair when the bow tension is high. However, the tension is not fixed; bows are stored in an un-tensioned state and manually re-tensioned for each practice or performance session. In addition, the hair and stick both noticeably respond to the environment so that tension may further change without human intervention.

Due to the shape of the bow, with the bow off the string, hair-stick distance is monotonically



(a) no co-linear points

(b) co-linear points only

Figure 4.3: Simplified demonstration of the 'displacement triangle' with two anchors A1 and A2, and three locations P1-P3 in (a) but only P2 and P3 in (b). A line has been drawn from each anchor through each point.

related to bow tension: hair-stick distance increases as bow tension increases. Indeed, string players commonly use hair-stick distance to estimate bow tension when re-tensioning. Conveniently, using optical sensors to measure the displacement triangle means we can also easily use them to measure the off-the-string hair-stick distance and obtain means for automatic tracking of bow tension.

One other major bow performance parameter is tilt (see Figure 2.3). Though tilt may slightly effect hair reflectivity by moving hair perpendicular to the bow's primary performance axis, in this work, tilt was largely ignored with sensors oriented to face the hair when bowed with an average tilt.

Work on estimating bow position and pressure based on optical sensors and the idea of a unique displacement triangle involved multiple revisions to both hardware and mathematical algorithms for estimating bow position and pressure. This chapter will start with the a discussion of the different sensors trialled, a discussion of both how to use them, and basic information regarding pros and cons of using each sensor. Discussion of sensors will also include differing styles of sensor mounting as mounting turned out to be a major practical consideration. The chapter will then continue with the derivation of algorithms for estimating bow position and pressure based on the measurements provided by the sensors including discussion of intermediate results. Finally, we will discuss end results and the overall usefulness of reflectance sensors bow tracking.

4.3 Using Near-Field Optical Reflectance Sensors

We performed two major iterations of hardware design using near-field optical sensors for bow tracking. The first design uses analog reflectance sensors mounted on either small hard circuits or flexible printed circuits, while the second design switched to integrated digital reflectance sensors on thin flat circuit boards. Both designs use four optical sensors and can be incorporated into the same mathematical approach to estimating position using the ideas behind the displacement triangle.

4.3.1 Optical Sensing in Musical Contexts

Optical sensing in musical instruments is quite an old idea. Wayne Stahnke used a pair of optical sensors to detect key press and velocity in the original Bösendorfer predecessor to the Disklavier, a design that remains in use today [44]. Middle distance infrared reflectance systems have also been in commercial use for a while with examples such as the Alesis AirFx [178] and the Roland D-Beam [147]. Examples of near-field reflective optical sensors are much more sparse. The Moog Piano Bar, which mounts on top of the piano keyboard, uses near-field reflection off the keys to detect key press [112, 111], while Leroy [95] used them to in an attempt to create an optical pick-up.

4.3.2 Analog Reflectance Sensors

The first design was based on analog reflectance sensors using LED light emitters and either a photo-transistor or photo-diode receiver with both transmitter and receiver sharing a common operational center wavelength. The reflectance sensor were chosen for being small, cheap, accurate and easy to work with. Four pairs of sensors were placed at different locations along the bow to capture the displacement triangle of Figure 4.2.

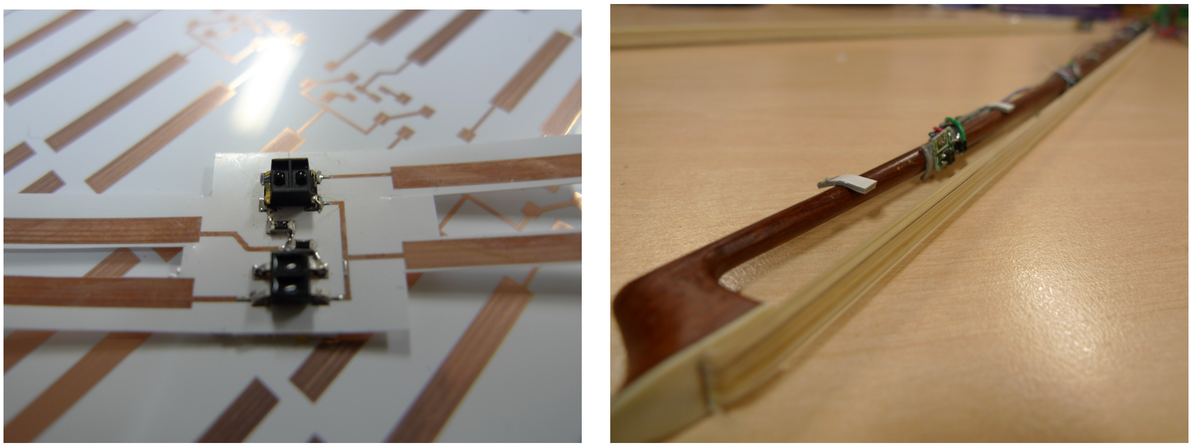


Figure 4.4: GP2S700HCP and QRE1113 mounted to flex circuit and on test bow.

Operational Circuitry

Figure 4.5 outlines the primary two circuit configurations for use with both the photo-transistor and the photo-diode. Current will flow in proportion to the amount of light received at the receiver and this can be measured using a load resistor R_L in either pull-up configuration or with an opamp. Using an opamp increases circuity, but reduces the impedance driven by the transistor output allowing faster operation. In either configuration the principle remains the same: $\Delta V_o = -\Delta I_o R_L$ where I_o is the current output of the photo-transistor and V_o is the measured voltage output. The more current flowing, the lower the measured voltage. The measurable range is otherwise determined by the choice of R_L and the supply voltage.

We investigated three physically low profile sensors: the Fairchild QRE1113, a reflective proximity sensor with photo-transistor output, the Sharp GP2S700HCP, also a reflective proximity sensor with photo-transistor output, and the Avago HSDL-9100, a proximity sensor with photo-diode output. Compared to photo-diodes like the HSDL-9100, photo-transistors produce more output current for a given LED transmitter current, but are slower. All three of these sensors are inexpensive and under 2.5mm high. The QRE1113 is 1.7mm high and has an optimal sensing range around 1mm. The GP2S700HCP is 2mm high and has an optimal sensing range around 3mm. The HSDL-9100 is 2.4mm high with an optimal sensing range around 5mm.

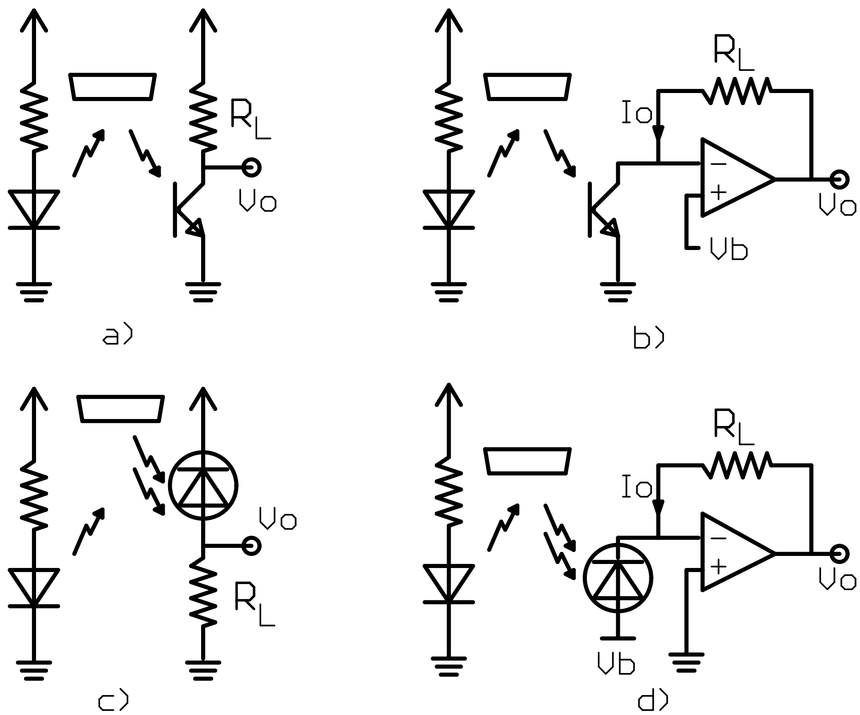
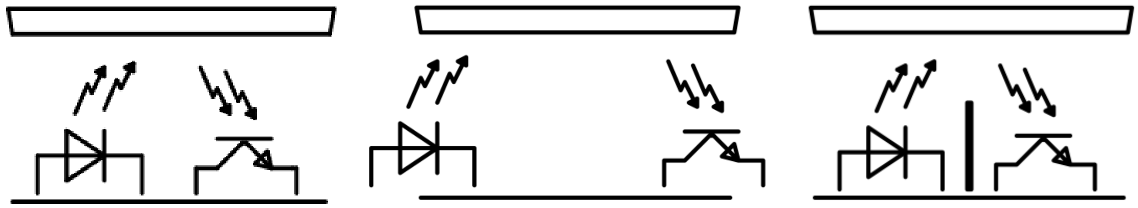


Figure 4.5: Circuit layouts for using photo-transistors with (a) a pull-up resistor and (b) opamp, or using (c) a photo-diode with a bias resistor or (d) opamp.

Characterizing Distance Response

Near-field optical reflective sensors are fairly easy to understand and demonstrate differences at the micrometer level, but absolute interpretation of near-field sensing is more of a challenge. There are a number of non-linearities and behavioral characteristics that must be assessed.

One of the key considerations in selecting an optical sensor is its optimal sensing distance and response curve. Optimal sensing distance, the distance just above the maximum current flow where response changes rapidly, is largely determined by design attributes such as the distance between the transmitter and receiver, the transmitter's angular radiation profile, both



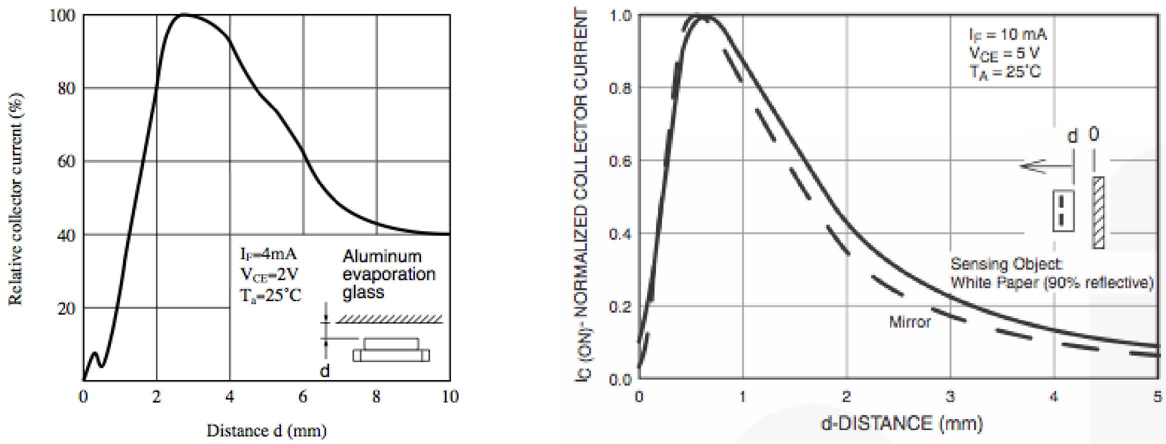
(a) Basic (b) Horizontal Variance (c) Altered Radiation Profile

Figure 4.6: Design variations effecting response characteristics of near-field reflectance sensors. Using (a) as a starting configuration, optimal sensing range can be altered by increasing the distance between the transmit and receive (b), or changing the radiation profile of the transmitter and/or receiver for instance through design of the LED or an artificial boundary between the two (c).

illustrated in Figure 4.6, or the receiver's responsivity. As an object comes closer, more of the transmitted light is reflected back towards the receiver increasing current flow. However, as shown in the accompanying data sheet graphs, when an object is too close, the reflected light from the LED may not fall within the sensor's viewing angle causing a decrease in current flow. Figure 4.7 provides the datasheet sample response curves for both the QRE1113 and the GP2S700HCP. It should be readily apparent that the curve is not only irregular, but that due to current drop off below the optimal sensing range, there is potential for ambiguity when interpreting results as there are two possible distances producing a specific current flow.

Another major consideration in near-field optical reflectance is the reflectivity of the object being measured. If the object is non-reflective, the current output and range of the photosensor will be dramatically reduced. Bow hair is sufficiently reflective to provide reasonable response.

The third major issue affecting the response curve is the current through the transmitter LED. While this does not significantly impact the optimal sensing distance, it will increase the sensing range, lengthening the response curve above the optimal sensing distance.



(a) GP2S700HPC (b) QRE1113

Figure 4.7: Voltage output response curves for GP2S700HPC (a) with an optimal sensing distance of 3mm and QRE1113 (b) with an optimal sensing distance of 1mm. *Taken from Sharp GP2S700HCP Datasheet, Oct. 2005, and Fairchild QRE1113 Datasheet, Aug. 2011*

Variation Within A Part Number

Early on it was noticed that different sensors of the same make yielded significantly different current response even when in the same configurations. With the QRE1113 sensors, we have used load resistors ranging between $12k\Omega$ to $47k\Omega$ to achieve similar range. We were even able to notice two physically different manufacturing styles for the QRE1113, each having noticeably different current output. The GP2S700HCP also suffers from this variability but to a lesser extreme. In a case where the test bow was exposed to an environment where the ambient light wavelengths significantly overlapped with the GP2S700HCP operating wavelengths, all GP2S700HCP sensors responded similarly, so we suspect the variation between sensors results from inconsistencies in LED strength. We have since worked in other projects with the Omron EE-SY1200 which, though more than three times more expensive than the other two, appears to offer far more stable performance across individual sensors.

Because of the variability between sensors, it was useful to determine a sensor's expected performance. A test jig (Figure 4.8) was built for both the QRE1113 and the GP2S700HCP that

Sensor	$10k\Omega$ pull-up	$20k\Omega$ pull-up	$1M\Omega$ pull-up	Opamp
GP2S700HCP	2.4kHz	800Hz	-	9kHz
QRE1113	1.4kHz	400Hz	-	9kHz
HSDL-9100	-	-	5kHz	22kHz

Table 4.1: Transient response for two photo-transistor optical reflectance sensors.

allows characterization of the sensor output prior to soldering into a circuit. The jig remains solder free by using compression to force contact with the test circuit pads. Removable acrylic plates enable test heights every 3mm. By testing the sensor before soldering, appropriate load resistance R_L can be selected to match the target operating range and appropriate transmit LED current can be verified.

Transient Response

In order to test sensor's transient response, each sensor LED was driven using a square wave while tracking the current output in both pull-up resistor and opamp output configurations (see Figure 4.4). The frequency of the square wave was then increased until the output transient failed to settle in time for accurate measurement. In general, when using the pull-up configuration, the transistor or diode has to drive a much higher impedance then when using the opamp which has an ideal input impedance of zero. Higher impedance reduces maximum operational frequency. As seen in Table 4.3.2, with a $10k\Omega$ pull-up, the QRE1113 was able to run at 1.4kHz, and the GP2S700HCP was able to achieve 2.4kHz. Switching to a $20k\Omega$ pull-up increased response range but slowed the QRE1113 to 400Hz and the GP2S700HCP to 800Hz. Using an op amp configuration, the frequency range was increased to roughly 9kHz for both the QRE1113 and the GP2S700HCP. All tests were conducted driving the sensors with 10mA and the results are consistent with the product datasheets.

According to the product datasheet, the HSDL-9100 has a rise time of only 6 μs, which would enable running at over 150kHz with a 5.1kΩ load. However, unlike the photo-transistor, the photo-diode only produces a very small current so that running at 3.3V, a more appropriate load resistance is at least 300kΩ. Driving the sensor LED with 100mA, and a $1M\Omega$ resistor at the output, it was able to run at 5kHz. Switching to an opamp configuration significantly improved the frequency response to 22kHz and possibly higher but doing so requires significant analog filtering to remove ring and noise from ambient light.

Analog Sensors in Bow Tracking Context

We started by placing four pairs of near-field optical sensors on the stick of an intermediate level wooden bow placed to maximize expectation of valuable data and distribution across usable stick space. We used pairs of sensors in order to accurately detect the full expected range of distances. For our configuration, the tensioned hair determined our maximum target detection range. Under tension, the hair-stick distance ranged up to 14mm although once mounting and sensor height are taken into account, the needed detection range is reduced to 0 - 11mm. The QRE1113 offers optimal sensing range down to 1mm making it a clear option, however it was not possible to choose an R_L that would provide sufficient response above 7mm without saturating the output voltage at lower heights. The GP2S700HCP has a wider range but empirical tests found current rapidly starts dropping below 2.5mm so many results would be ambiguous as to which side of the 2.5mm maxima they were on. The HSDL-9100 demonstrated detection out to 20mm with an output maximum at 4.5mm. As the QRE1113 is optimal under 4mm, a range overlapping with the GP2S700HCP's optimal range, the combination of the two provides a usable solution.

The approximate locations of sensor pairs are represented in Figure 4.2. The bottom 'frog' sensors were placed 75mm from the frog with the 'lower' bow sensors 165 mm from the frog. The 'tip' sensors were mounted 70mm from the tip with 'upper' bow sensors 190 mm from the tip, 580mm and 460mm from the frog respectively. No sensors were mounted at the extrema as there is little measurable deflection where the hair is secured to the stick, nor did we mount any sensors on roughly the middle third of the stick. During performance, there is typically minimal or even no clearance between the hair and stick in the middle section of the bow. A sensor placed too close to the center may not just contact the hair, but far more problematically, clip the string as the sensor moves past.

The clearance issue places a significant constraint on the selection of sensors. If a sensor is too large, it will not be safe to mount places where meaningful data can be collected. All three sensors were chosen in part due to their low profile. Clearance issues also restrict the amount of circuitry as thick or large circuit boards are more likely to catch a string.

Dealing with Non-Linearity and Range

Due to the non-linear curvature of the voltage vs. distance output response curves shown in Figure 4.7, variability between sensors, and sensitivity to test environment, it is necessary to calculate the output voltage curves empirically. To control for physical environmental factors, the sensors needed to be tested in the 'as used' configuration.

Figure 4.8: The various test tools (clockwise from rear)- the Razer Hydra and accompanying sensor, scale for future pressure tests with a height jig, height jig used for determining height lookup table, solder-free QRE1113 test jig for sensor evaluation.

A voltage vs. distance curve was derived by tests performed using the actual test bow. It was important to test with the actual planned bow setup as bow hair has different width, density, and reflectivity at each end. The QRE1113 and GP2S700HCP were co-located at points along the bow as previously described. Another test jig (Figure 4.8) was built to separate the

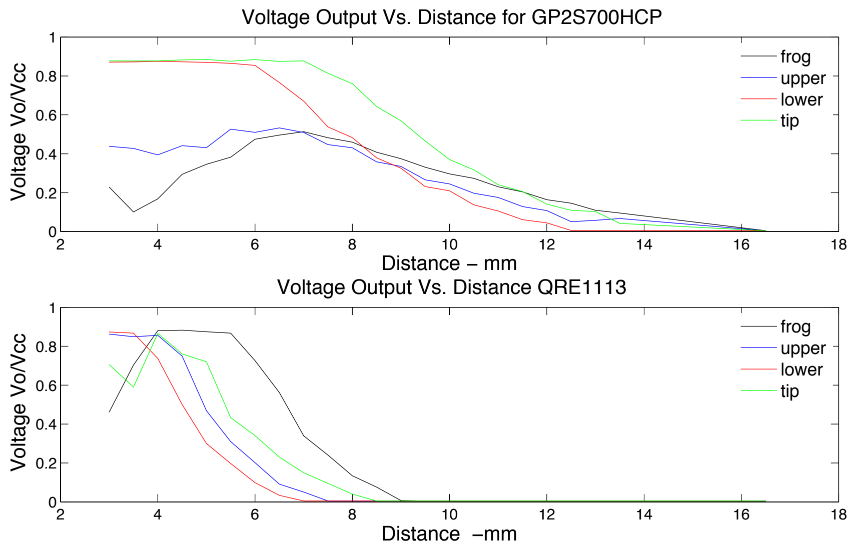


Figure 4.9: Test voltage vs. distance response curves for the GP2S700HCP (top) and the QRE1113 mounted on a violin bow.

hair and the stick at every 0.5mm for distances between 3 and 13.5mm from the stick. Test values taken using the jig were used to build a height lookup table. Figure 4.9 illustrates the various voltage vs. distance curves. Results are largely in keeping with the expected curve shapes in Figure 4.7. An important result is that for each sensor pair, the detection region for the QRE1113 must begin before the peak sensing output of the GP2S700HCP. This is key to resolving the ambiguity introduced by the drop of the response curve after the the peak output.

Using the height lookup table, it is now possible to estimate displacement heights. We are used linear interpolation between data points. By translating the voltage readings from both sensors to height based on empirical data, we are reducing some of the non-linearities of the optical sensors. With the data now in the same reference frame of mm, we can also combine the results from the QRE1113 and the GP2S700HCP into a single distance estimate. We used a fairly simple algorithm: first, if the distance is out of the QRE1113 range, we use the distance measured using the GP2S700HCP. If the GP2S700HCP estimate is higher than the QRE1113, we assume the GP2S700HCP is operating closer than its optimal sensing distance and rely entirely on the QRE1113. If the QRE1113 height is higher than the GP2S700HCP, we take a weighted average. Figure 4.10 illustrates the unified height estimates for a test sample where the GP2S700HCP can be seen providing the long range distance estimate with the QRE1113 providing the close range estimate when the GP2S700HCP bottoms out.

Drawbacks of Analog Sensors

Though analog near-field reflective sensors are cheap, small, fast (in the context of tracking human motion) and use simple circuits, the variability between sensors and need to use two sensors to cover the expected detection distances make their use complex. Furthermore, noise was a major issue. Although the receiver's optimal response wavelength is matched to the transmit wavelength, it is possible to have interference from the environment. Although all three sensors transmit around a 950nm wavelength, we did not see any evidence of interference between the sensors considered. Amplifying the small μA current output for the HSDL-9100 also amplified a small 50Hz frequency noise within the system power supply, highlighting the low signal-to-noise ratio.

Under fluorescent lighting, neither the QRE1113 nor the HSDL-9100 appeared to suffer from noise due to ambient light, however the GP2S700HCP response frequencies did overlap with more common lighting wavelengths. Fluorescent lighting introduced a small 100Hz $50mV$ hum leading to roughly 0.25mm of error. While error from fluorescent lights was marginal, as all three types of sensor operate in the IR range, ambient light in IR rich environments such as direct sunlight, incandescent light, flash lamps, the Kinect, or as commonly found in theater lighting may cause moderate to substantial interference. Severe interference was obvious when switching from testing in an office environment to a recording studio where there was such a substantial bandwidth overlap that studio lighting saturated some sensors.

If practical, noise can be dealt with to some degree by directing the sensors away from noise sources, taking a baseline of ambient noise, or increasing transmitter current which allows reducing receiver sensitivity without loss of range. The hum from florescence seen in the GP2S700HCP was sufficiently specific that it should be removable through a notch filter. However, for our target of building a tracking system practical for general use, the sensors can not be hidden from ambient light and must function in a variety of uncontrolled environments

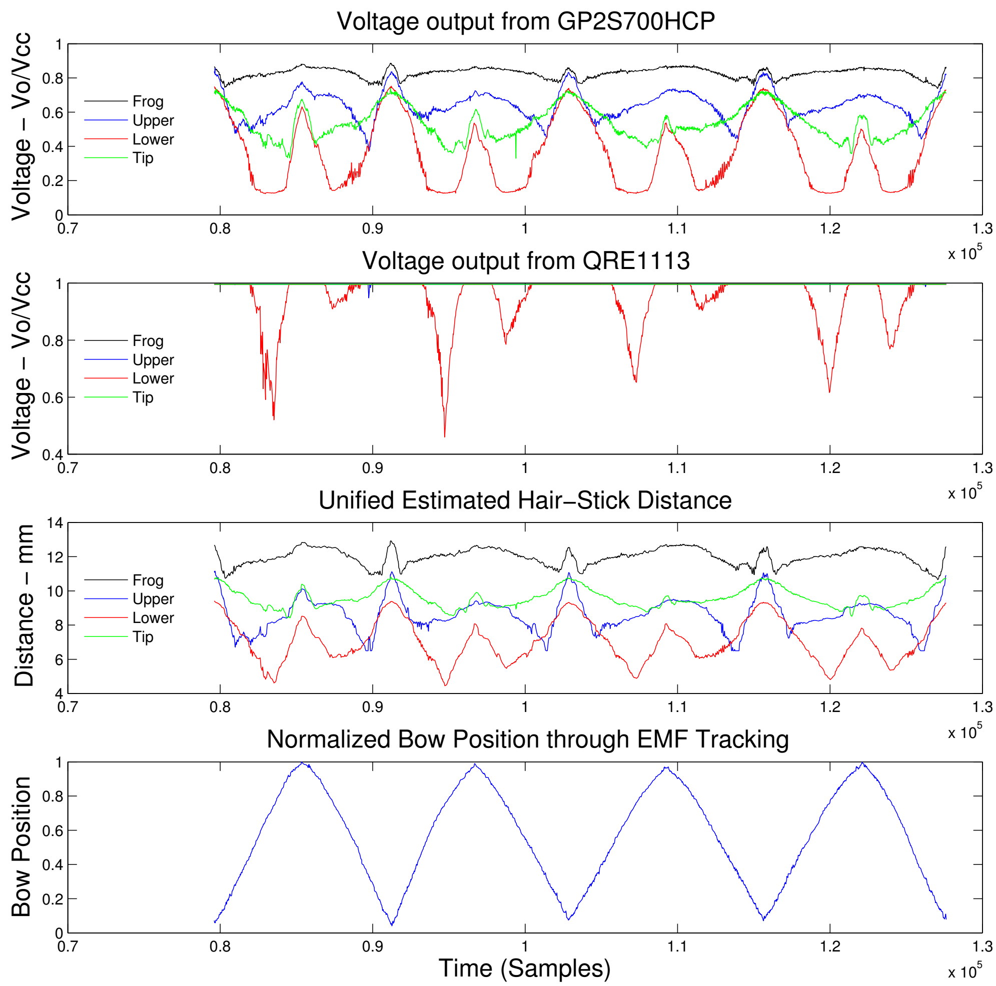


Figure 4.10: Data taken from a sample test of long bows played from the frog to the tip. From top to bottom: GP2S700HCP voltage, QRE1113 voltage, combined distance measurement, and ground truth bow position as estimated by the Hydra.

including stage performance. As such noise using analog sensors was too problematic and for later builds we switched to a digital optical sensor.

4.3.3 Digital Optical Proximity Sensors

Digital optical proximity sensors use the same basic principle as analog sensors, measuring reflected IR light, but drastically reduce noise interference by modulating the transmit IR light at frequencies over 50kHz. A reasonably affordable low-profile digital optical proximity sensor that almost completely eliminates ambient IR light noise is the Vishay VCNL4000, providing 16-bit resolution over an effective range of 1-200mm, with optimum performance under 5mm.. Tests of the effects of ambient IR light on the Vishay VCNL4000 found nominal variation in the baseline reading when there was no measurable object, but no variation when measuring the presence of a reflective object. A secondary benefit of the VCNL4000 is that the useful detection range covers the full range of expected hair heights meaning rather than two sensors per measurement location, we only need one. With only one sensor per measurement location, there is no need to perform the height calibration described in Section 4.3.2 necessary to merge the two analog sensor ranges into a unified distance estimate. As a result, we switched to the VCNL4000 in later designs.

Information to and from the VCNL4000 is passed using I^2C. Through I^2C it is possible to program the transmit LED current, modulation frequency, and proximity detection sample rate. The VCNL4000 can be set for automatic position sensing which has a maximum rate of 250Hz or 4ms per sample, or can measure by request. When manually requesting proximity, using an 100kHz I^2C clock rate, the VCNL4000 can sample distances at 712Hz, every 1.4ms.

However a significant drawback of the requirement to communicate to the VCNL4000 through I^2C is that it only has one I^2C address so that only one chip can be on the I^2C bus talking to our primary hardware controller, an Atmel AVR32UC3C, at a time. In order to use four VCNL4000s, we paired each with an ATTINY45 programmed to control a switch passing or blocking the data line to the VCNL4000 based on either two direct addressing lines or I^2C communication with the ATTINY45. Adding the ATTINY45 and switch require more than doubling the size of the board required to support each measurement location on the bow.

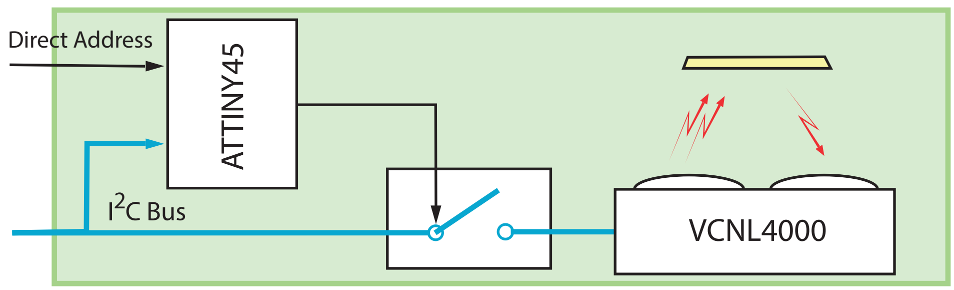


Figure 4.11: On-bow digital sensor circuitry: a VCNL4000 digital optical reflectance proximity sensor with access to the I^2C bus controlled by a local ATTINY45. The ATTINY45 opens and closes the gate depending on either direct addressing or instructions received over I^2C.

This arrangement required six wires off the bow. By using direct addressing and interleaving commands, it is possible to sample all four measurement locations within 3.4ms letting us run bow sampling at 294Hz. Using I^2C for addressing takes an extra 500us per sensor.

4.3.4 Physical Construction, Mounting & Cabling

Space, weight, and comfort being key design considerations, we put significant effort into physical manufacturing details. Sensors towards the middle of the bow needed to have minimal height to reduce the likelihood of bow hair pressing directly against them and catching on the string. Circuit boards also needed to be narrow to minimise how much they protrude past the bow stick. As players frequently tilt the bow during performance, protruding edges can catch a string. In our trials, catching the edge of the circuit board was the most common serious interruption to regular play. Sensors, circuits, wiring, and mounting need to be light to minimize the change in how the bow feels. Mounting methods also need to be removable without damaging the bow but stable enough to keep sensors in place.

We initially used a combination of analog sensors mounted on small 10x10x1.6mm circuit boards and board-less sensors with wire-wrapped leads taped to a wooden bow. The width of the boards meant catching was a problem and attaching a flat board to a curved surface meant sensor rotation could vary. In order to improve on both issues, we switched to printed flex circuits.

Printed conductive flex circuits allowed for the inclusion of large flexible contact leads that were wrapped around the bow providing substantial surface area for taping the sensors to the bow. Areas with inflexible chips were backed by thin cardboard so the plastic circuit substrate did not separate from the soldered chip when the flex circuit was bent around the bow. While this method worked well for securing the flex circuit and effectively minimised the circuit and sensor profile, we found the circuit traces broke too easily. Wrapping the flex circuit around the bow exceeded the effective bend radius making circuit traces delicate. They would break beyond repair when knocked or if a connecting wire was strained. Although bulkier and not as easy to stably mount, we found it necessary to return to using thin hard circuit boards for the revised build.

The third build of bow sensors used the digital sensing circuitry described in Section 4.3.3 using 36.6x9.1x0.8mm circuit boards. These boards were made as narrow as possible and placement was carefully done to minimize the likelihood of catching. Though catching did still occur with this build, it was significantly better than the first rigid circuit board experiments. Secure mounting still remains somewhat of a challenge. Tape still allowed change to board tilt. Instrument putty provided the best instrument safe mounting alternative and meant we could add height to place sensors at the frog or tip much closer to the hair. While this improved measurement signal to noise ratio at those two locations, the putty is too flexible for stable mounting. The last build uses tape and balsa wood mounts to provide height at the tip and frog and hot glue behind boards to block rotation. While we recognize hot glue would not be acceptable on a high quality wood bow, the experimental bow used is carbon fibre making hot glue an expedient solution.

Wiring & Cabling

Wiring should be minimal to minimise distraction, likelihood of catching during regular use and transport, and also limit the size and weight of cabling off the end of the bow. Wiring from the frog off the bow also needs to be mounted in a way so that it does not interfere with regular play. We intentionally opted to cable the bow rather than use wireless connections as the addition of batteries to support wireless systems inevitably impacts balance. Still, the cabled approach requires supporting the cable. Cabling is routed along the forearm using straps so cable weight and pull is not felt at the bow.

Initial trials using analog sensors ran 10 long cables directly off the bow that were then strapped to the forearm. Though expedient for prototyping, due to the cabling along the forearm, it was non-trivial to pick up the bow or set down the bow. To improve this, a magnetic connector near the wrist was added to allow easy disconnection of the bow. The player wears three straps, two at the wrist to support the cable near the bow, and one farther up the arm so that the cable doesn't flap with fast hand gestures. The two wrist straps sit either side of the connectors, with the bow side strap providing optional velcro support between the bow and the bow-side connector.

4.4 Polynomial Fitting to Estimate Bow Position

With the basic idea of the 'displacement triangle' and results from the four sensing locations, it is easy to visually interpret the data to estimate bow position. Returning to Figure 4.10 and looking at the unified estimated hair-stick distance, at either end of the bow there will be very little overall deflection due to the tension near the anchor points. Starting from the frog, as the bow contact point moves towards the first sensor, the deflection height at that sensor will rapidly drop to its minimum. The estimated bow hair height detected at the other three sensors will also drop but at a slower rates depending on how far they are from the frog. As the string passes the first sensor, that sensor hits its minimum and the height to the hair will now increase the farther the bow travels. In the meantime, the second sensor is approaching its minimum and so on. After passing the last sensor at the tip, the measured distance will increase for all sensors as deflection decreases and the next bow stroke back towards the frog begins. Pressure can be estimated by the total scale of the deflection across all sensors.

While it is theoretically possible to combine the calculated relative height displacement with a physical model to estimate bow position and pressure, just as the sensors suffer from multiple non-linearities, so does bow behavior. It is not just the bow hair that flexes, but the bow itself. Flex and stiffness vary significantly between bows so while we can intuitively assess rough bow position and pressure from the results, more adaptive methods are required for useful accuracy. We did initial position only investigations into the feasibility of optical sensors for bow tracking before expanding to a more useful model including pressure and tension.

4.4.1 Initial Investigation into Bow Transversal Position

For our initial investigation into whether optical reflectance sensing was feasible for bow tracking, we limited investigations to position tracking only. We decided to use multi-variate polynomial regression to extrapolate position. In our case, the measured bow heights are the four independent variables, and bow position is the conditionally dependent variable for fitting. Since the relationship between the measurements and position should be stable for each sensor arrangement and bow, once we have a polynomial describing that relationship we can use it to extrapolate further bow positions.

In order to capture a dependent variable 'ground truth,' we used the Razer Hydra[1] from Sixense. As an EMF tracking system, tracking using the Hydra follows the same principles described by Maestre in [101]. We used the first generation Razer Hydra which has a more limited range (∼75cm) and less accuracy (∼2mm) than more expensive systems, but at under $100 it is dramatically more affordable. The Razer Hydra was designed as a hand-held game controller which means the full controller is far too large to act as a bow tracking method on its own. Convenient use of the Hydra requires some hacking in order to reduce the size to the EMF sensors and circuitry only and in comparison to the bow, the EMF sensors are large (20mm) and too cumbersome for practical performance use[2] Still, the Hydra is sufficient for

[1] `http://sixense.com/razerhydrapage`

[2] Using EMF for bow tracking in performance requires tracking both the bow position and the violin position so apart from comfort issues, using a Razer Hydra for performance tracking would require mounting a second

rough estimate of position in training.

One EMF tracking sensor was mounted to the tip of the frog while the author, an experienced violinist, played a number of bow strokes using the test bow with the optical sensors. At the beginning of each test, the bow was placed at the frog and tip a number of times in order to define the bow's vector of motion. In order to ensure minimal lateral variation, there was a fixed target for right hand extension and minimal movement otherwise. 'Ground truth' bow position was derived by translating and rotating the EMF data from the vector of motion onto a single axis (Figure 4.10). Tests were run using a variety of stroke pressures to cover the full range of potential hair position.

A convenient result of our approach is that once the polynomial fit is found, it should remain valid for a bow provided the same sensors and placements are used. Although the approach does require calibration, we expect it should be a one time calibration per bow after which the Hydra is unnecessary.

Test & Results

A sample test result is displayed in Figure 4.12. Bow position was normalized so that the full length of the bow, frog to tip, is a range of zero to one, and measurements from the analog sensor pairs were unified as in Section 4.3.2. A polynomial transform was determined using *polyfitn*,[3] a Matlab package by John D'Errico. Polynomial of best fit was derived using three data sets totalling 7033 samples at 100Hz to fit the four dimensional set of optical sensor heights to normalized bow position. The total number of bow strokes in the fitting set was 29: 15 down bows and 14 up. Each set of data was composed of full tip to frog bow strokes of varying pressure: one heavy, one medium, one light. The polynomial was then used to estimate bow position based on a data set of hair-stick distances not included in the fitting set. We tested up to a sixth degree polynomial using root mean squared error to evaluate performance.

The sample result set in Figure 4.12 clearly follows the correct bowing pattern. Using a fourth

set of sensors on the violin.

[3] `http://uk.mathworks.com/matlabcentral/fileexchange/34765-polyfitn`

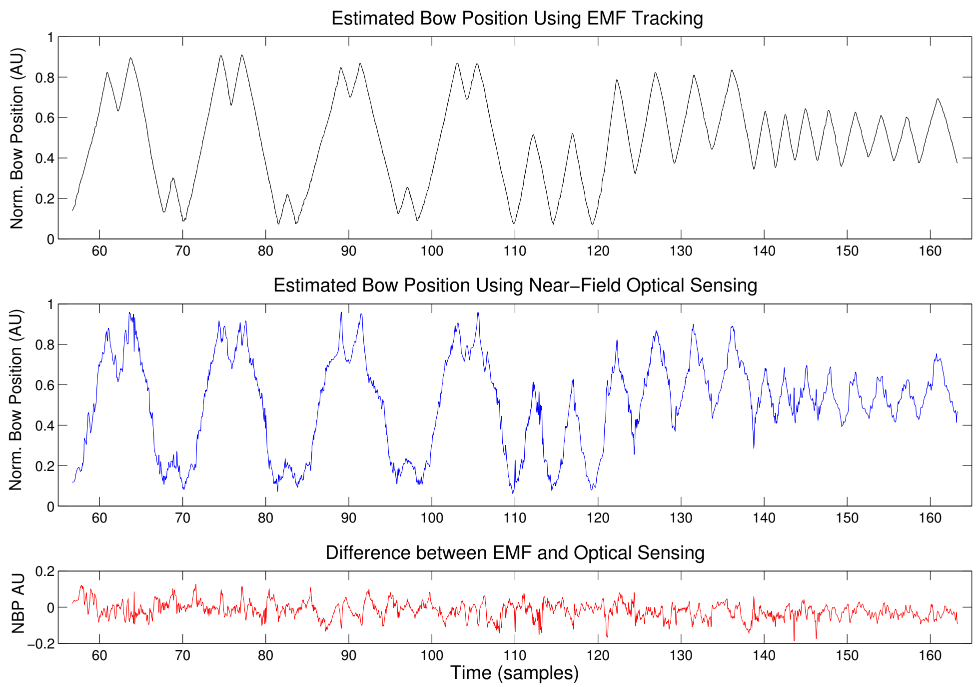


Figure 4.12: Sample result demonstrating mixed bow strokes. Top: bow position as recorded through EMF tracking. Bottom: estimate of bow position based on near-field optical reflective sensors.

degree polynomial derived from the three training sets, the test set was 3468 samples long and when compared to the normalized EMF result had an RMSE of 0.0635AU (Arbitrary Units[4]). The raw error was never more that 20%. Expanding the fitting data to include a data set with continuously changing bow pressure reduced the RMSE to 0.051AU. Similar tests rotating target and training data sets suggested a typical RMSE of between 0.04AU and 0.09AU and that the fourth order polynomial was generally optimal. These results were all generated prior to any filtering method which would be expected to improve performance. For instance, smoothing results across an 8 sample window reduced the RMSE of the first data set from 0.0635AU to 0.0574AU.

Conclusions

Our preliminary results using near-field optical sensors on the difficult task of bow position tracking demonstrated the potential for near-field optical reflective sensors as a powerful means for fine distance measurement and capturing person-instrument interaction. Though the preliminary analog design, build, and polynomial fitting was intended as a rough first attempt, there was obvious success from which to build upon and continue work including the addition of measuring bow pressure.

4.5 Estimating Bow Position, Pressure, and Tension

Having demonstrated the basic feasibility of bow position tracking using analog near-field optical sensors, the next step was adding tracking of pressure, tension and investigating improvements to methodology. The addition of pressure and tension to position means we now have three dependent variables which we must describe through our four independent variables, the proximity measurements along the bow. Continuing with our non-linear treatment of the system, we deal with the added complexity by using our trained polynomial functions to estimate expected sensor measurement for each bow position, pressure, and tension. Then,

[4] As used in this paper, units are not really arbitrary as 1AU is equivalent to the length of the bow hair, 650mm. However normalized bow position more usefully aligns with how violinists describe bow position.

given a set bow tension for a test session, we attempt to find which combination of position and pressure is most likely to yield the measured proximity values.

Methods presented in this section were developed using both analog and digital sensors builds, however all data and results in this section were collected using the digital proximity sensors and circuitry described in Section 4.3.3 as the digital sensors proved far more resilient to noise in ambient light. The VCNL4000 provides us with a single height estimate for each sensor location and, as proximity readings are only ever used in polynomial fitting, we use raw readings without converting to metric distance measurements.

4.5.1 Deriving Polynomial Estimates of Expected Sensor Results

Revisiting the model of the 'displacement triangle' from Figure 4.2, it is expected that the hair-stick distance measured at each point along the bow is determined by the the tension, τ, the location where the string touches the bow, l, and the pressure of the bow against the string, ρ. From our preliminary investigation, it was clear that as a sensor passes the string the slope of the measurement function effectively inverts, an event that can not be described using a single polynomial function so we switched to describing proximity as two polynomials either side of the sensor. We model the hair-stick distance function, h_k, at a given measurement location as a continuous piecewise function with two pieces, g_k above, and h_k below, the point l_k, where $k \in 1...4$ denotes the specific sensor on the bow. Using four sensors, we end up with the following four piecewise functions that describe the sensed distances for a point in time t:

$$h_k(t) = \begin{cases} f_k(\tau(t), l(t), \rho(t)) & \text{for } l < l_k \\ g_k(\tau(t), l(t), \rho(t)) & \text{for } l \geq l_k \end{cases} \Bigg|_{k\in\{1,2,3,4\}} \tag{4.1}$$

While the actual form of f_k, g_k may be unknown, we assume that reasonable polynomial approximations f'_k, g'_k for these functions exist. We derive f'_k, g'_k using empirical training data from recorded bow strokes, capturing bow tension, position, and pressure using the setup shown in Figure 4.13.

Collecting Training Data

Refining the previous means for bow training drawing straight strokes on a violin, for the revised design, we switched to moving the bow on a mock string (a short, thin wire) which rests atop a USB scale recording pressure. Similar to the preliminary trial, a Razer Hydra electromagnetic tracker is attached to the frog of the bow to record bow transverse position. Training bow strokes, executed by hand, cover the full bow length and a range of realistic pressures, and were accomplished using a standard bow grip. Due to the scale having a low 5Hz sample rate, all strokes and pressure changes were slow.

Once a training set has been collected, a bow tension metric for the set is calculated, bow tension altered, and a new training set collected. Tension has an important effect on the sensor readings and therefore on the equations for calculating bow position; however, the actual tension measurement in Newtons is never needed, so we calculate a simple tension metric as the average of the hair-stick distance of all four sensors, sampled when the bow is off the string and horizontal. It is expected that bow tension will remain stable through any moderate length performance, an assumption backed up by results from the training sets. There was typically less than 0.05 mm difference between the tension metric at the start and end of each set.

Figure 4.13: Picture of prototype bow (left) and training setup (right): Razer Hydra and scale with mock string.

During training, bow transverse location is determined by finding the location of the attached electromagnetic position sensor with the test string at the frog and then at the tip. Provided each training stroke uses the full bow length, the distance between the tip and frog measurements should always be the bow's length. This allows the frog-tip vector to be used to rotate results into a single dimension, assuming that the bow always moves along a straight vector. This assumption, sensor noise, and human error resulted in the expectation of as much as 5% error in the position training data as determined by comparing electromagnetic position estimate with known fixed locations on the bow. The scale data is also shifted in time slightly to minimize the effects of the low sample rate and latency of the scale. Additionally, the mock string and slow bow speed used in training strokes (required for the slow digital scales) result in excessive bouncing of the bow on the test string. This bouncing is reflected in the optical sensor measurements, but is not captured in the slower scale measurement readings, further introducing training measurement error.

[64] describes an effective bow pressure calibration technique building on [155] that uses similar principles with more accurate sensing technologies. However, in keeping with our low cost target, we intentionally limited ourselves to simple, cheap, and portable off-the shelf technologies accepting that this will reduce the accuracy of our result. We also found that for our calibration tools, the continuous bowed strokes were both easier to accomplish, and, by spanning the entire bow to produce far more data points, yielded better results than obtaining sensor estimates at a pre-defined set of combined locations and pressures.

Polynomial Curve Estimation

Now we can do the actual curve estimation for the functions given in Equation 4.1. Again, we use *polyfitn* in Matlab to perform three dimensional linear regression. The sample size was largely determined by the need to run enough sample sets to avoid overfitting in the tension dimension [9], provided a reasonable range of pressure and position values were visited each for tension, led by default to a reasonable fit in other dimensions. For results presented here, twenty training sets were produced, each running 2-3 minutes using the full bow length with downward forces ranging 0.15N - 4.0N, producing over 130,000 data points. As tension, τ, was treated as constant for each set, with 20 points of comparison, the polynomial fit order of τ was restricted to the second or third degree to reduce the risk of overfitting along that dimension.

Evaluation of the curve of best fit was done through cross-validation using a one-set-out approach. Initial experiments suggested that overfitting became observable with 5th-order or higher polynomials. As a result, polynomial orders above 8 were not evaluated. For each of the eight polynomial equations in Equation 4.5.1, the optimal order was considered independently and chosen by finding the order that yielded the lowest average root-mean-square error (RMSE) across all test sets. Figure 4.14 illustrates sample training and test sets. Any discontinuity within $h_k(t)$ between polynomials f_i' and g_i' is smoothed by the use of a Gaussian weighted average of the two centered around the sensor location. Subsequent results demonstrate that the curve fitting method sometimes gets temporarily stuck at the transition point between the two curve estimates for a given sensor set. This suggests that the curves are non-continuous and a different method of transitioning between curves of best fit may be beneficial.

4.5.2 Using Curves to Estimate Bow Pressure and Position

Having found the polynomial curves to calculate each expected optical reading for given tension, position and pressure, we can estimate the expected set of hair-stick distances for any real-world bow state. Outside of the test environment, these hair-stick distances are the only measurements we have to infer information during a performance. Again, the bow tension metric is the average of hair-stick distances with the bow off the string. We assume τ is fixed and determine τ for a session by measuring the hair-stick distances with the bow off the string just prior to playing.

Given the tension, τ_θ, for the session θ, we generate the expected measurement curve surface, a two-dimensional matrix – with position l along the x-dimension, and pressure ρ along the y-dimension – for each (l, ρ) pair for each sensor. Figure 4.15 provides an example of the expected resulting surface for a given sensor and tension. Assuming ideal curve fits, a given hair-stick distance measurement h_t will determine the slice of the curve set which is the possible (l, ρ) pairs that would produce that result. Due to expected noise, we assume the actual (l, ρ) pair should be near, but not necessarily on the ideal slice.

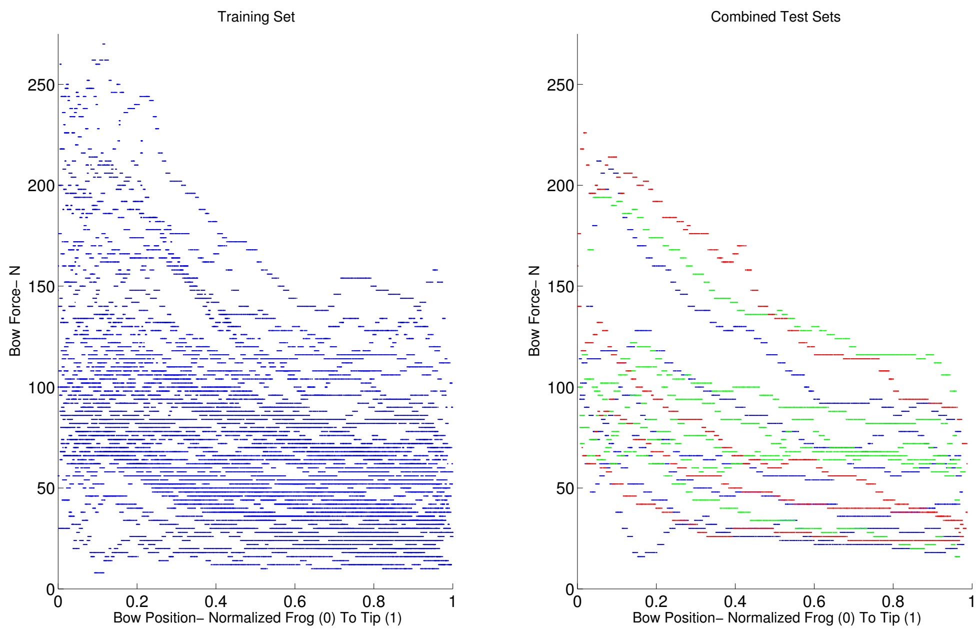


Figure 4.14: Sample data sets for training (left) and testing (right). Three different test sets were used in testing. Test sets were included in the training sets when not being tested against.

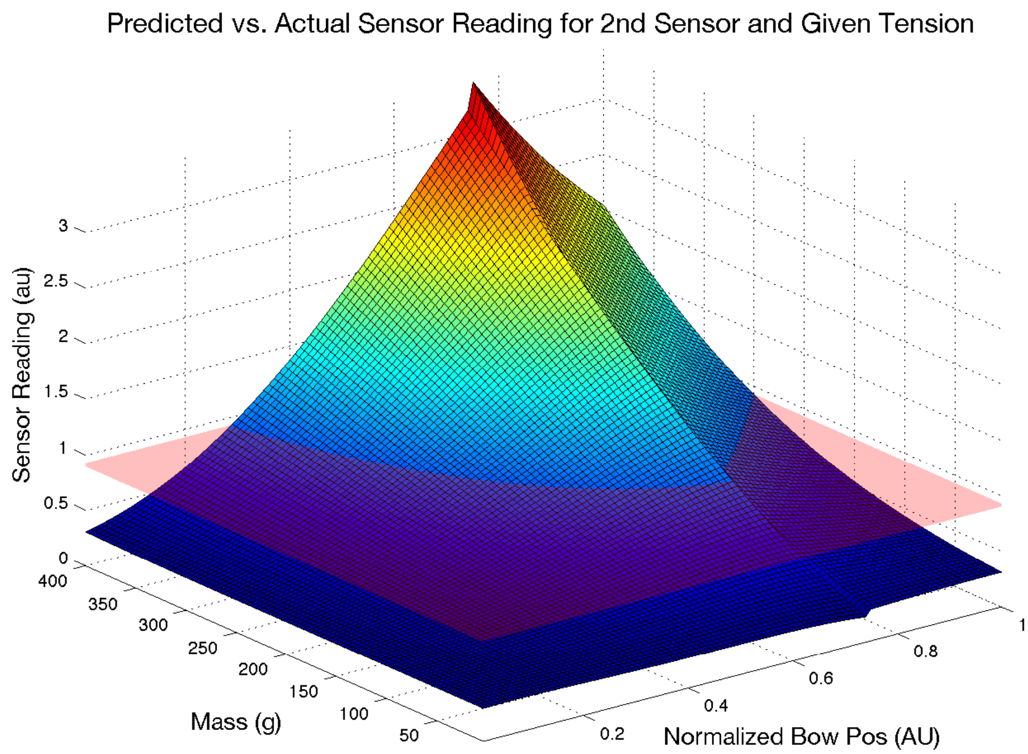


Figure 4.15: Sample curve surface showing the expected reading for a sensor at a given tension. A sensor reading defines z-axis level for the horizontal plane slicing through the curve surface with the resulting slice defining the set of possible position and pressure combinations that would result in that reading.

During performance, we compute the absolute difference between the measured result and each sensor's matrix of expected possible results. The estimate difference is scaled by the quality of the corresponding curve fit, as given by inverse of the fit's RMSE which was calculated while determining best polynomial fit:

$$d(l, \rho, t) = \sum_{k=1}^{4} \left| \frac{h_k(t) - f'_k(\tau_\theta, l, \rho)}{\text{RMSE}_k} \right| \tag{4.2}$$

The RMSE also implicitly reflects the overall variation of a given sensor's measurement range so this scaling also balances the different sensor's contributions. This scaled difference between expected and measured is then summed across all sets. As illustrated in Figure 4.16, with τ_θ fixed, the (l, ρ) pair for which the expected distances across all four sensor locations provides the lowest summed difference is theoretically the actual momentary location and pressure:

$$(l(t), \rho(t)) = \arg\min_{l,\rho} d(l, \rho, t) \tag{4.3}$$

4.5.3 Utilizing Expected Time Series Continuity

Effectiveness of bow position and force estimation was evaluated by cross-validation using the same one-set-out method used for curve selection. The average RMSE for the estimated vs. actual normalized bow position was 0.165AU and the average RMSE for the estimated vs. actual force is 0.296N. We can improve results by noting that bow motion and downward force should be continuous, i.e. any estimate should neighbor the previous estimate.

Estimates are weighted using a Gaussian with the mean, μ, centered on the previous estimate at $t - 1$, so that neighboring points are more likely to be selected as the optimum. This approach has the drawback that when wrong, the estimate may get stuck in local minima. This issue is partially alleviated by changing the weighting distribution based on the likelihood of the last estimate's correctness. If the previous estimate is believed to be incorrect, the weighting is flattened by altering the Gaussian's deviation σ, so that a distant point may be considered equally likely.

Confidence (using the colloquial meaning of the term) is initially assessed using the distance of the previous estimate, $d(l_{t-1}, \rho_{t-1}, t-1)$ (Equation 4.2). With ideal data and ideal equations

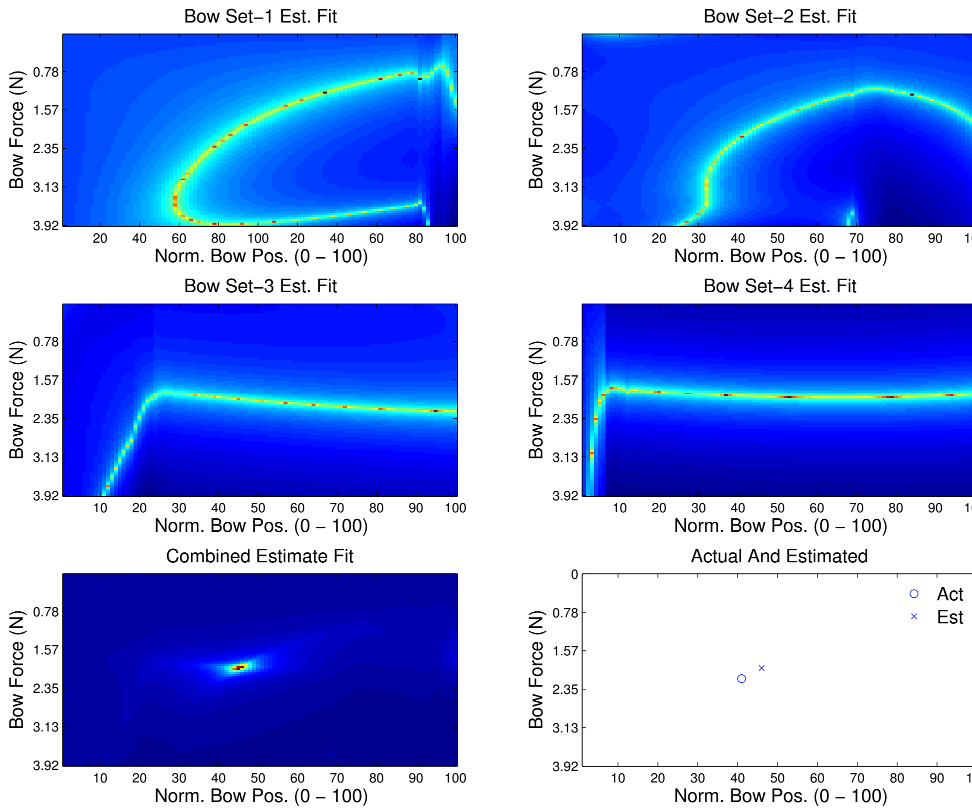


Figure 4.16: Distance between expected hair-stick distance and actual momentary hair-stick distance for all reasonable (l, ρ) pairs for each sensor (top four, clockwise from upper left: frog, lower bow, upper bow, tip) the sensor sets combined (bottom left). The minimal distance metric for $d(l, \rho, t)$, is chosen as the expected bow location $l(t)$ and bow pressure $\rho(t)$, shown vs the actual measured position and force (bottom right).

describing the relationship between sensor reading, position, force, and tension, $d(l_t, \rho_t, t)$ would be zero. The non-zero difference, $d(l_t, \rho_t, t)$, that does occur is attributable to real-world error and becomes a useful estimate for correctness.

Experimental results suggest some additional trends in error that do not correlate with the basic confidence factor, $d(l_t, \rho_t, t)$. First, if estimates suggest the bow is not moving, although it could reflect actual performance, it is more often evidence of the nearest neighbor requirement causing the estimate to get stuck. In this case the confidence is decreased, to reduce the effect of nearest neighbor weighting allowing the algorithm to jump to a better but more distant bow estimate. Second, the position and force estimation algorithm is most likely to be wrong and get stuck at lower forces, often resulting in the estimate for position at the extreme tip or frog. As a result, we decrease the confidence for estimates with low pressure or at the extrema of the bow. We have also tried using an unscented Kalman filter to optimize results and balance measurement error with state, but it did not outperform the flexible weighting design.

4.5.4 Estimation Effectiveness

Weighting to select nearest neighbors decreased the RMSE for the estimated vs actual normalized bow position from 0.165AU to 0.121AU and the average RMSE for the estimated vs actual force from 0.296N to 0.260N. Generally the weighted neighbor restriction significantly improved results when the force is above 0.6N. In fact, within the training set, for forces that measured above 0.6N, the weight of a standard violin bow (61g), the RMSE for the normalized position estimate, 0.063AU, suggests an expected error of 6%, which is only marginally above the expected 5% accuracy of the electro-magnetically measured bow transverse position. Similarly, force estimates tend to be worse near the frog and excluding the bottom 10% of the bow, which is rarely used in practice, reduces the force estimate RMSE to 0.190N. To give pressure results context, the bulk of regular play ranges in force from 0.5N to 2N. Sample results are shown in Figure 4.17.

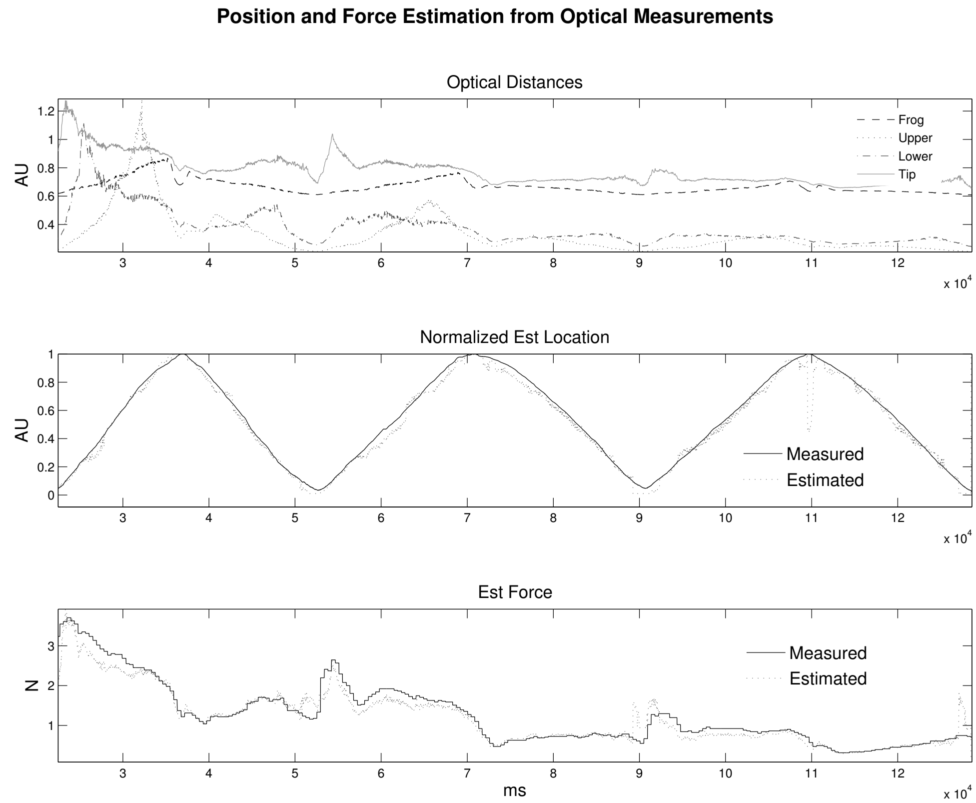


Figure 4.17: Raw sensor readings, bow position and downward force estimates from a test set with measured position and measured force. The sensor readings are taken at the four sensor locations on the bow and are used to derive from the hair-stick distance measurements in the top graph.

4.5.5 Further Experiments

We experimented both with sensor placement and varying the number of sensors. The results presented here were generated with sensors placed at 45mm, 151mm, 485mm, and 624mm from the frog. The bow hair is 650mm long. The positions targeted measuring at bow extrema and equal distribution along the length of the bow while not interfering with bow use.

Training was done with two further sensors at 530 mm and 581mm from the frog and controlling VCNL4000 addressing through I^2C. Using all six sensors did improve results but not dramatically. Based on a subset of the full training set, using six sensors resulted in a position estimate RMSE of 0.114AU compared to 0.121AU and a force estimate RMSE of 0.256N compared to 0.260N. We felt that the additional latency required to poll six sensors (9.6ms) outweighed the 5% performance improvement. We also experimented with using either of the two added sensors instead of the sensors at 624mm or 485mm, or using only three sensors but found all of these reduced accuracy with little or no added benefit.

Additionally we experimented with building training data by capturing bow deflection at set positions and weights. While this method was successful at yielding polynomial fits, without building or buying a specialized test rig, the calibration process was overly time-consuming and yielded far fewer training data points. If we were able to find a low-cost digital scale with higher sample rate, we could presumably improve our calibration process significantly both in accuracy and speed.

4.6 Discussion

Within the dominant range of violin play– the top three quarters of the bow and the full bow weight on the string– our estimates are useful. We expect results for both position and pressure to be correct within $\pm 3.8cm$ and $\pm 18.6g$ respectively. Section 6 demonstrates various types of note onset successfully detected using our methods for bow tracking. Though absolute error may be higher than preferred, differential behavior is reasonably accurate especially for pressure estimates. As can be seen in Figure 4.17, both position and pressure typically follow the correct basic trend making data useful when looking for behavioral change.

Beyond note onset, such information can be useful in providing insight into a performers action. For instance, a common difficultly teaching is to convey differences in bow weight on the string as this is altered by subtle shifts in muscle support throughout the right arm and hand. Being able to demonstrate a rough estimate of applied pressure along with accurate performance trends could help provide the additional insight needed to successfully illustrate the task.

It is expected that beyond the outermost sensors correct estimation of position and pressure will be error prone as sensor inputs should be theoretically monotonic with respect to proximity of the sensor. As we need two non-linear points to determine the location of the 'displacement triangle' apex, measurements at the very tip or frog break this requirement. In these areas, hair deflection is also very small due to the proximity to the hair anchors further increasing the likelihood for error. Error due to the difficulty of interpreting minimal displacements at bow extrema is illustrated in the rightmost low pressure bow-stroke Figure 4.17, where at the very tip, the position estimate suffers from a major error.

Still pressure estimate accuracy appears to be stable everywhere apart from the frog. Bow position is less consistent. Though highly accurate in the middle of the bow, due to difficultly estimating at the tip or frog, position estimates suffers error at the tip at low pressure. The other main area the estimate gets stuck is when passing the two inner sensors. Though only limited evidence of distortion when passing sensors is visible in Figure 4.17, it is more visible in the result estimates depicted in Section 6. Though there is smoothing between the two polynomial curves for a given sensor, transitioning between the two is clearly still a source of error.

Keeping our bow tracking low cost and portable was also successful. The total cost of the analog sensors outfitted to bow was under $10 with digital sensors under $15. Even including the cost of the analog or digital circuitry, the supporting processor, and the Hydra, the cost remained under $150. This is not only easily 30 times cheaper than most bow tracking methods, but also keeps the cost within a reasonable consumer range. It is also easily portable. It requires the sensor equipped bow, an Arduino-sized control board, a laptop or other means of conveying data to the user, and associated cabling.

4.6.1 Importance of Sensor Stability

Crucially, the accuracy of the bow estimates is highly dependent on physically stable sensors; even a slight change in sensor angle or position can dramatically impact estimate accuracy. Because tension, used to calculate curve estimations, is derived from a combination of all four sensors, error from a single sensor propagates to the estimates from all sensors. If the original position can not restored, it requires retraining the bow. For instance, the first build used analog sensors without circuit boards that were taped directly to the bow and unlikely to move. Though crudely built, it yielded some of our best results (typical results were within 0.04 - 0.09AU). In build versions relying on tape, small knocks and resting the bow on the sensors caused them to rotate and shift slightly but the tape did not easily allow the subtle readjustment to return the sensors to previous positions leading to the need to retrain.

Results in this section used the slim digital sensor circuit boards attached using a combination of double sided tape and instrument putty which was instrument safe but malleable. The bow setup used for the results discussed here did experience multiple knocks during transportation and use which disturbed the sensors. Though finding a better way to fix sensors remains desirable, a reasonable level of consistency was achieved by using the ratios of raw measurements between the different sensors during training as tuning references. If the off-the-string readings from a particular sensor noticeably differed from expectations, the bow would be tensioned so that the other sensors matched previous test sets and the offending sensor would be gently nudged till the readings for the perturbed sensor roughly matched historical readings. This method was not only highly practical, but it also established a typically average set of off-the-string readings and demonstrated the importance of sensor physical consistency. Sets deviating farther from this rough average returned at least two to three times worse accuracy for position and force estimates.

4.6.2 Bow Feel and User Reaction

The various iterations of bow build have weighed in between 73g and 76g, 13-16g heavier than the un-augmented violin bow (59-61g) but only slightly more than a viola bow (68-72g). Improvements in the manufacturing process would allow us to reduce the weight, and also improve accuracy and feel. For instance, we are using unnecessarily heavy wire to connect to the sensors which could be better replaced with lightweight small connectors. Also, finding a suitable chemically neutral tape or other means of safely stiffly securing the sensors would enable use of optical sensors with professional quality bows.

The bow was user tested as part of the 8 person pilot user test discussed in depth in Chapter 8. The build in that study used balsa wood supports (lighter and more stable than putty) at the tip and frog and, as the bows we worked with in initial user tests were carbon fibre, used hot glue and double sided tape to secure the sensors. Users rated the bow heavy but acceptable. Only one player stated the bow had altered balance with all others saying the augmentations only had minimal effect on performance technique. The cable was deemed more intrusive, though the test case was ill-served by the length of cable between the bow and the magnetic connector which was initially too short. The bow has also been used by the first author in public performance of pieces that are technically demanding for the bow hand without adverse affects.

An intriguing immediate result was the clarity with which bow bounce was revealed. Every bow has a point where it will naturally tend to bounce even in a long applied bow stroke. Similar bounce can happen in an uneven bow change. Bounce results in uneven sound, something rarely desirable and whose removal is a common practice target. Optical tracking easily captured this bounce. The software for data collection includes a real-time display of incoming data. While playing, the author was able to see the bounce, data that in post-processing might look like noise, and attempt to react to minimize it. While typically aware of bounce through feel and sound, it was informative to have it visually highlighted. For further details on user reactions to the bow, see Section 8.4.10

It is worth noting that with the sensor positioning in this section, it is possible for the string to catch the sensor placed 151mm from the frog. Likelihood of catching the string is directly affected by bow tension, and how much the player tilts the bow. Manufacturing improvements should resolve this issue in future.

4.6.3 Further Applications of Near-Field Optical Sensing on the Bow

While not attempted in this thesis, the micrometer resolution of near-field optical sensors should enable the capture of bow tilt. Despite tilt being fairly ubiquitous in playing, with the exception of Schoonderwaldt [153], tilt tracking is fairly neglected. The primary purpose of tilt is to affect tone by controlling the width of hair in contact with the string. Hair in contact with the string will be deflected inward at an angle to the stick while hair not in contact with the string remains largely in its normal tensioned location. Co-locating two sensors, one directed at the hair and one slightly angled should provide information on tilt based on differences between the two reflectance measurements. Only two sensors are required as string players tilt almost exclusively away from the bridge. Alternatively, some more expensive digital proximity sensors include 'gesture' sensing using multiple IR receivers angled to capture light reflected from different directions.

Additionally, the capture of bow flex, measurable by comparing polynomial sets for different bows, is in itself, a useful result. It seems reasonable that it is possible to derive relative stiffness at different points along the bow as the multiple sensors capture relationships between deflection at these points. Bow deflection characteristics are not directly linked to audible results, but are crucial when considering bow quality and feel for the player.[4]

4.7 Conclusion

We have demonstrated effective low-cost and portable real-time tracking of bow position, pressure, and tension using near-field optical sensors and the idea of the 'displacement triangle'. We showed how to use both analog and digital sensors in bow tracking applications along with issues involved in either choice. While both are inexpensive, small, and provide high resolution for close proximities, digital reflectance sensors have superior noise rejection making them the better choice for our needs. We also demonstrated how to deal with complex bow mechanics by using polynomial fitting to estimate expected sensor readings for given position, pressure, and tension. Once polynomials have been trained, we can compare measured to expected sensor estimates across all sensors and, placed in the context of previous estimates, estimate current position and pressure. Using materials costing under \$150, and under normal performance conditions with full weight of the bow on the string and excluding the bottom 10%, our experiments found bow position tracking had an average RMSE error of 0.063AU and pressure estimates had an RMSE of 0.19N.

Tracking estimates can be used in a multitude of situations. In Section 6 we use bow tracking estimates to find performance note onsets from off-string attack, bow change and slurred repetition, while we have also found raw data from sensors visually highlight bow bounce. It could also provide other useful real-time information about performance of bow tasks during lessons or practice. Beyond traditional performance, tracking estimates can be used artistically as the real-time nature provides opportunity for real-time illumination of performer actions. For instance, as described in Section 7.3 we have used both raw data and pressure estimates to control both computer graphics and stage lighting.

Though we have proved it is possible to perform low-cost portable bow tracking, it could use further improvements in mounting and calibration which would make training last longer and be easier to carry out. Stable mounting in particular has proved a major factor in accuracy and would benefit from improved solutions. Additionally, improvements to the estimation algorithm especially around sensor locations would help improve results and eliminate estimates becoming stuck. Finally, basic improvements to mechanical build would help bow tracking to be more robust and less intrusive to normal performance.

Chapter 5

Low-Latency Audio Pitch Tracking: a Multi-Modal Sensor-Assisted Approach

This chapter incorporates significant material from 'Low latency Audio Pitch Tracking: A Multi-Modal Sensor-Assisted Approach' by Pardue, Nian, Harte, and McPherson, originally published in the proceedings for NIME 2014 [133] and 'A Low-Cost Real-Time Tracking System for Violin' by Pardue, Harte, and McPherson, published in JNMR 44.4 [134].

5.1 Introduction

Our experiments into assisting pitch learning through real-time intonation feedback using both aural and visual methods (Chapter 9), and potential pitch simplification (Chapter 8), require a pitch correction system that meets three primary criteria: 1) it must be accurate enough for users to trust the system, 2) using pitch estimates for pitch correction does not cause distracting audio glitches, and similarly, 3) the delay in obtaining the pitch estimate is fast enough not to be a disturbance. Monophonic pitch tracking is sometimes considered a solved problem in audio analysis, but existing approaches do not always meet the stringent accuracy and timing demands of live instrumental performance. Musicians can detect reaction latencies of 20-30ms in musical instruments [106, 1], and latency under 10ms is an accepted target for interactive audio systems [72, 49]. Expert live performance also demands an extremely low detection error rate.

Most audio pitch tracking algorithms are based on windowing the signal and performing analysis using fast fourier transform (FFT) based spectral analysis or direct computation in the time domain, such as in autocorrelation-based approaches. But a commonly-used window size of 4096 samples at 44.1kHz requires 93ms to fill; even a 512-sample window, short enough to degrade the performance of most algorithms, lasts over 11ms. Where multiple consecutive frames are compared to improve robustness, this will multiply the total latency. In many algorithms, minimum window size is inversely related to the lowest frequency to be detected, reflecting inherent tradeoffs of time and frequency resolution in short-time spectral analysis. Audio-based pitch estimation is potentially more accurate for offline analysis but it fails to perform consistently for real-time latencies under 10ms.

Sensors measuring the performer's actions can instead be used to detect pitch [60], but the mechanics of most instruments makes these solutions incomplete. Pitch on a violin is determined by the location of fingers pressing the strings and the open string tuning which depends on how the violin is tuned and also varies over time, sometimes needing retuning within a single performance. Capturing only performer action fails to account for potential variation in string tuning.

We used a hybrid approach which uses sensors to produce a fast, rough initial pitch estimate combined with an audio estimate that is not dependent on knowing the string tuning in advance. The estimate restricts the search space of a low latency audio analysis algorithm whose accuracy would otherwise be unacceptably low. Because audio and sensors tend to produce different types of errors (harmonics being common in audio, variations in calibration with sensors), the combination can be both fast and accurate. This chapter demonstrates two approaches to position sensing on a violin fingerboard, with sensor-assisted versions of two commonly-used audio analysis algorithms.

5.2 Audio Only Pitch Detection

Pitch determination within a monophonic recording can be performed quite reliably off-line [34, 171], but it is more difficult in real time, especially when targeting less than 10ms latency. There are many potential approaches to monophonic pitch tracking.

5.2.1 Monophonic Pitch Tracking and Autocorrelation

Violin pitch tracking falls into the general problem of monophonic pitch detection. In-depth reviews of monophonic pitch tracking are presented in [6, p.16-20] and [33] (which focuses on voice intonation but includes many non-voice specific algorithms). There are three primary means of pitch detection: spectral methods, temporal methods, and combinations of the two. Spectrally-based methods work by short-time Fourier analysis of the audio signal, looking for a dominant frequency and its harmonics. The dominant frequency can be identified in several ways, including autocorrelation within the spectral domain [93], cepstral analysis [123], and constant-Q methods [17].

Temporal methods are commonly based on autocorrelation. By cross-correlating a windowed signal with itself, peaks occur in the autocorrelation function at harmonics of the fundamental frequency within that window [141]. Ideally, the largest peak is the target frequency; however, resonances and variations in the signal result in non-trivial error rates. Since basic autocorrelation weights results equally no matter the lag, it often leads to the frequency estimate being too low. An alternative is to reduce the weight at greater lags by effectively reducing the correlation window using the equation for biased autocorrelation:

$$r_t(\tau) = \sum_{j=t+1}^{t+W-\tau} x_j x_{j+\tau} \tag{5.1}$$

While Equation 5.1 reduces the likelihood of underestimating the fundamental frequency, it does this at the cost of increasing the likelihood of overestimating the fundamental as shorter periods are weighted more heavily. A third variation of temporal-based autocorrelation is to use a normalized difference function:

$$d_t(\tau) = \sum_{j=1}^{W}(x_j - x_{j+\tau}^2) \tag{5.2}$$

Kawahara and Cheveigné's Yin [34] is based on this difference function, Equation 5.2, which reduces the influence of amplitude changes on the autocorrelation function to achieve a dramatic decrease in errors. Yin has been proven to be highly robust and reliable and has become the dominant means for pitch detection.

5.3 Fingerboard Tracking: Capturing Finger Placement

We track finger placement by configuring the violin fingerboard to act as a linear position sensor; an approach similar to those explored by both [60] and [51]. We opted to pursue a resistive approach as though capacitive methods for position tracking have clear potential [62], flexible circuits currently mountable on the violin can be delicate, and require precise regular calibration, complicating usability and reliability. Similarly, inexpensive video tracking approaches also suffer from limited accuracy, and are difficult to calibrate [68, 35]. As we look to keep the sensor low cost, robust, and accessible, we seek to improve upon resistive based linear position sensors by exploring new sensing arrangements that are non-intrusive and instrument-safe while retaining seamless fingerboard feel and appearance. We performed significant design tests of two builds, the first based on using the actual violin strings as a conductive part of the circuit, and the second, more similar to a standard linear potentiometer where conductive strips were placed under the string instead.

5.3.1 First Design Using Strings as Conductors

A typical means for making a linear potentiometer strip is to place a conductive layer over a resistive strip with an air gap separating the two. When pressed, the conductive material contacts the resistive strips at the point of pressure, forming a voltage divider. Commercial linear potentiometers have a built-in air gap which requires minimal, but still noticeable, height, and with that, noticeable edges. The height and edges alter the feel of the instrument, making commercial sensors non-ideal for use on traditional instruments.

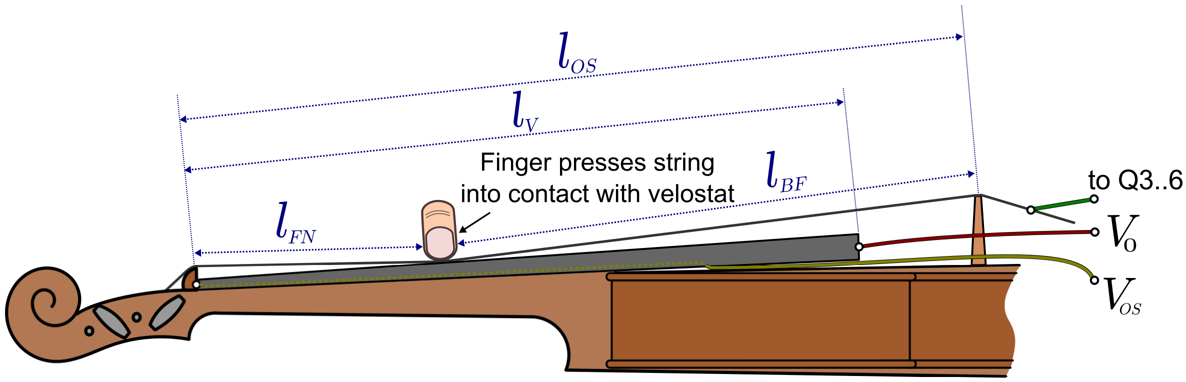


Figure 5.1: Fingerboard sensor configuration. The fingerboard is covered with a layer of resistive velostat. Contact between a string and the fingerboard produces an electrical connection. Q3 to Q6 determine which string is detected and are controlled by Figure 5.2.

In our first attempt at building a seamless fingerboard sensor, we applied a bare resistive strip directly to the fingerboard, making use of the conductivity of metal strings and the normal air gap between the string and the fingerboard (Figure 5.1). A single thin layer of *velostat* (carbon-infused polymer) was glued to the fingerboard with connections hidden under the bridge end of the fingerboard. Along with its resistive properties, the velostat has a smooth finish similar to the fingerboard. Each violin string was electrified by running the ball end through an electrically-connected solder tab. We removed the fine tuner from the violin E-string as it was made of non-conductive plastic. We successfully used both generic inexpensive strings and high-end professional strings. With the exception of added wires from the fingerboard, the feel of the violin is only minimally changed.

Though this physical arrangement felt natural and did not significantly change the feel of the violin, experiments made it apparent that successfully fingering a note does not require the string to touch the fingerboard. While less of a problem with physically larger strings, the finger itself acts as the stop mechanism altering pitch, not the fingerboard. This led to frequently missed finger presses, especially on the E-string. As a result, we reapproached the design of the sensor to capture finger to fingerboard contact, rather than string to fingerboard contact.

Additionally, due to the high resistance of the velostat, the conductive properties of the human hand introduces significant noise, reducing accuracy with single string contact to around $\pm 6mm$. Using a lower-resistance material would reduce the noise from human interaction substantially, but as we have quite high noise tolerance for left hand input, the velostat is acceptable.

5.3.2 Second Design Utilizing Custom Printed Linear Potentiometer

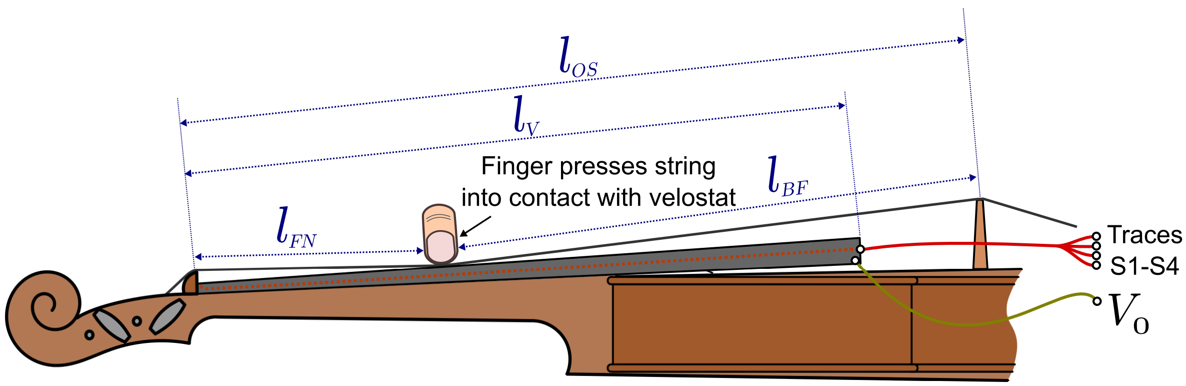


Figure 5.2: Fingerboard sensor configuration. The fingerboard is fitted with a custom resistive position sensor overlay. Pressing a string down with the finger causes an electrical connection to be made; the induced voltage being linearly related to the position of contact.

The fingerboard, as shown in Figure 5.2, is covered with the custom linear position sensor shown in Figure 5.3. The sensor comprises four printed conductive traces, one aligned under each string, spacer tape placed between the traces, and then a sheet of velostat placed on top. Finally, a thin laminate sheet is placed over the velostat to protect it from wear, dirt and oil from fingers. This arrangement is similar to commercial sensors but with the layers reversed. An air gap between the velostat and the conductive traces is created by placing thin tape between the traces prior to attaching the velostat. When a finger presses on a string, the velostat is pressed into contact with the conductive strip, forming an electrical circuit with the conductive trace below.

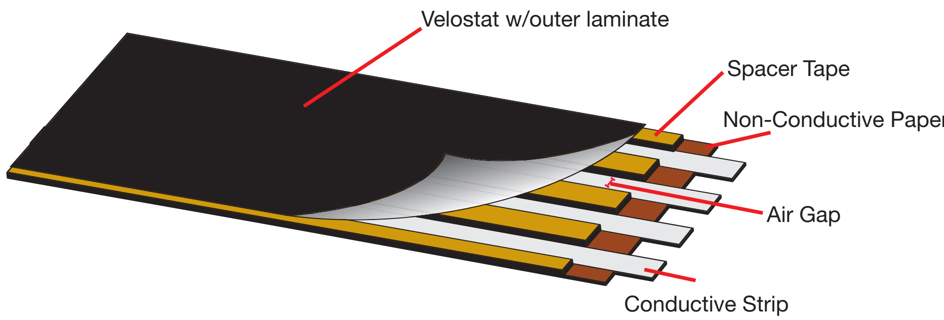


Figure 5.3: Fingerboard linear potentiometer design.

5.3.3 Capturing and Estimating Performed Pitch

For both designs, if the open-string length l_{OS} and length of the fingerboard velostat l_V is known then the finger position can be determined by driving the bridge end of the velostat with a fixed current and measuring the voltage v_{out} induced between the bridge end and the conductive means for a particular string. The ratio of v_{out} to the maximum measurable voltage v_{MS} (i.e. the voltage measured for an open string) can then be used to calculate l_{FN}, the distance of the finger from the nut:

$$l_{FN} = l_V \big(1 - \frac{v_{out}}{v_{MS}}\big) \tag{5.3}$$

From this, if we know the tuned pitch, f_{string}, of an open string then we can estimate the sounded pitch, $\hat{f}_{hw}$, of that string:

$$\hat{f}_{hw} = \frac{f_{string} \times l_{OS}}{l_{OS} - l_{FN}} \tag{5.4}$$

To find v_{out} we use the circuit given in Figure 5.4. A current source (here, a current mirror based on a BC212 PNP transistor) provides a constant current to the contact location dependent fingerboard resistance, producing an output voltage which is linearly related to resistance: $V_o = IR_C$. The resistor R_C should be chosen to be greater than the maximum

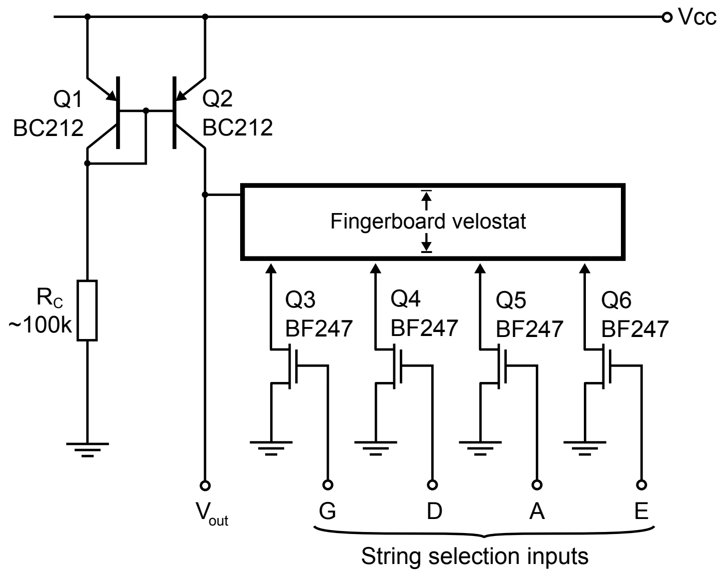


Figure 5.4: Circuit for fingerboard position detection. A current mirror supplies fixed current to the velostat making v_{out} linearly proportional to the distance traveled through the resistive velostat. A JFET, BJT, or other analog switch is used so that only one of the conductive traces passes current thus allowing measurement of finger position for each string separately.

resistance of the velostat strip in order to detect across the full length of the strip. Our strips had a typical resistance of around $80k\Omega$ so we chose $R_C = 100k\Omega$. In both designs we multiplex the velostat by grounding either a conducting string or a conductive trace in a repeating sequence, thereby allowing measurements to be made for all four strings.

The same custom Atmel AVR32UC3C board as in Section 4.3.3 is used to sample the voltage for each string and calculate a normalized position for the finger closest to the bridge. A sample rate of 300kHz is possible with each sample being the average of 4 ADC measurements.

Position accuracy was found to be $\pm 3mm$ leading to a rough estimate pitch value derived from Equation 5.4 accurate to within 16 cents on a stable, well-calibrated system. However a stringed instrument is not always well tuned, velostat conductivity is temperature sensitive, and a performer's body has varying electrical conductance. These all introduce variance into system performance.

Present builds of the fingerboard sensor use a tape (0.15mm thick double-sided polyimide) that leaves no residue, is strongly adhesive, and re-adheres so that the fingerboard sensor can be treated as a removable sticker. As each fingerboard is slightly different, once cut to size, the sensor can be easily placed (or removed). If the velostat surface wears, it is also possible to recut and replace just that part. After we found resistance of the velostat changed due to direct interaction with oils and friction from users, we counteracted this by protecting the contact surface with a spray plastic or other laminate.

As reviewed detail in Chapter 8, initial user trials with eight moderate to professional violinists proved the effectiveness and comfort of the sensor. All eight stated the sensor was unobtrusive and did not interfere with play, with five participants giving strong positive remarks as to its seamlessness. It has also been used successfully by the first author in multiple public performances featuring virtuosic repertoire. The thickness of the sensor means that it may not fully fit all fingerboards. A qualified luthier had to raise the nut to compensate on our cheap test instruments. The author has also performed with the sensor on an antique high-end instrument (over $20k) adding thin shims to effectively raise the nut held in place by string tension. This avoided recutting the nut with no ill effects.

5.4 Hardware Assisted Low latency Pitch Tracking

5.4.1 Sensors + Audio

Though widely used, autocorrelation methods, including Yin, typically yield results that, when wrong, are substantially numerically different from the correct frequency. A typical example is harmonic error, when many algorithms identify a harmonic of a frequency rather than the fundamental. Shorter correlation windows have lower inherent latency but are more susceptible to such errors. A trade-off between error and latency is unavoidable.

An additional issue is the fundamental limitation that autocorrelation methods require one or more periods of a wave to fit in the analysis window, placing a lower bound on window size. Correlation windows are unable to assess frequencies below f_s/W where W is the window size and f_s is the sample rate. The window must be at least as long as one period of the lowest frequency it can detect. For instance, the low G on a violin (196 Hz) takes 5.1ms for a complete period, and on a cello, the low C (65 Hz) takes 15.3ms for a complete period.

On the other hand, pitch errors from hardware sensors tend to be local inaccuracies, missing the correct pitch by a few percent in either direction depending on calibration and linearity. As such, combining hardware and audio approaches can yield useful improvements. Ajay Kapur demonstrated this concept with his *E-Sitar* [82]. The E-Sitar used electronic sensing to detect which fret is being played and audio analysis to compensate for the pitch variance from bending the strings.

With sensor-assisted pitch tracking, we begin with a rough estimate of pitch based on finger position. This allows us to use even the most basic pitch detection algorithm with a shorter analysis window to achieve an accurate result. Finger placement eliminates harmonic errors. Additionally, where the sensors indicate a fundamental frequency below what could be detected with a given window size, we can instead search for a harmonic of that frequency using a shorter window, allowing us to estimate the fundamental in shorter than one period.

5.4.2 Restricted Search Biased Autocorrelation

We used two audio methods for sensor-assisted pitch estimation: biased autocorrelation with quadratic interpolation, and a variation on Yin [34]. Biased autocorrelation, Equation 5.1, is commonly used in signal processing as it weights events with less lag more than higher-lag events thus creating a decay envelope. The decay envelope can be seen in the top of Figure 5.6 as the function flattens for high τ. Typically, estimated pitch period is determined by picking the highest peak in the autocorrelation function. However, this method is prone to error, especially at smaller window lengths when there are fewer accumulated wavelengths or notes below the minimum detection frequency. For instance, tested against a sample violin recording, using basic biased autocorrelation with a window size of W = 2048 samples had a 4.8% error rate, but a 2048 sample window incurs an unacceptably high 46ms latency. Reducing the window to 512 samples, the error rate rose to 21.1%. At 256 samples, corresponding to an inherent latency of 5.8ms at a 44.1 kHz sample rate, the error rate grew to 41.1%.

We combine sensors with audio by restricting the search for optimal pitch within the autocorrelation results to an area around the hardware estimate (see Figure 5.5), using the sensor-based pitch estimate from Equation 5.4 as the center point around which to search for the maximum of Equation 5.1. For autocorrelation-based pitch detection, errors tend to occur at harmonics of the correct pitch. While octave errors are most common, fifths are sometimes also found. With this in mind, we started by restricting the autocorrelation search space to be within a just-intonation major third (±25%) of the touch sensor estimate. Subsequent trials found that for small windows, where pitch estimates were less robust, it was useful to use a whole-tone search window (±12.5%), and using the original sensor design, hardware pitch estimate was typically within ±8% of the correct pitch. The switch to a custom printed linear potentiometer not only reduced missed fingers, but also improved typical accuracy to within ±2% or ±25 cents.

If there is no position estimate from the fingerboard, we assume the audio must be produced by an open string and restrict our search to a band of narrow frequencies around each string.

If the window length, W, is too short to effectively evaluate a low frequency, rather than search

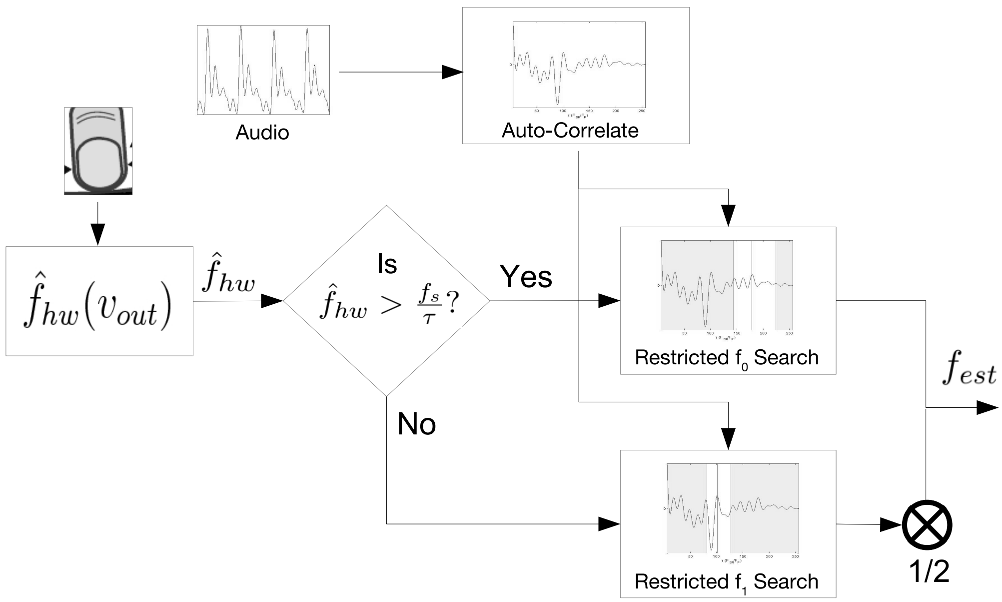


Figure 5.5: Hardware-assisted pitch tracking flow-chart. Pitch is estimated based on voltage signal from hardware using Equation 5.4 while autocorrelation is calculated on the audio signal. If the hardware pitch estimate is high enough to fall within the autocorrelation detection range, the algorithm searches for the best pitch within a whole tone of the hardware estimate. If the frequency falls below the detection range, the second harmonic is sought using autocorrelation and the result is divided by 2 to find the fundamental frequency.

for the fundamental frequency, we search the area around the expected second harmonic as demonstrated in the second example of Figure 5.6. The second harmonic will sometimes be a minimum instead of a maximum of the autocorrelation function, since the lag corresponding to the second harmonic also represents a 180° phase shift of the fundamental. We thus search for either a minimum or maximum in this case, choosing the stronger of the two responses.

Lastly, the autocorrelation function can only be evaluated at integer multiples τ of the sample interval, even though the actual period of the signal may lie between those multiples. In an effort to estimate the theoretical maxima, we use parabolic interpolation based on the peak value and the two surrounding points [34]. If the optimal lag τ is on the edge of the window and the correlation function is monotonic, we do not perform interpolation.

5.4.3 Restricted Search Yin

We have also implemented a frequency restricted search version of Yin [34], as Yin typically yields much better results than biased autocorrelation. Yin improves on simple autocorrelation in five stages. The first is the use of the Equation 5.2, the modified autocorrelation based on difference in signals rather than the raw signal. This method increases resistance to errors due to amplitude change. Second, Yin uses a cumulative mean normalized difference function (Equation 5.5) to reduce "too high errors", by weighting a result on its difference from a running cumulative average.

$$d'_t(\tau) = \begin{cases} 1, & \text{if } \tau = 0, \\ \frac{d_t\tau}{\frac{1}{\tau}\sum_{j=1}^{\tau} d_t(j)} & \text{otherwise} \end{cases} \tag{5.5}$$

We incorporate the sensor data restriction in the third stage of Yin. Since Yin starts from a difference-based autocorrelation variation, it looks for a minimum instead of a maximum and defines an arbitrary threshold which any result must be below. Yin selects the minimum from the first contiguous set of values under the chosen threshold, or the overall minimum if nothing is below the target threshold.

Rather than using an arbitrary (though effective) threshold, we replace the threshold search by the same frequency restriction technique used in biased autocorrelation, restricting the search

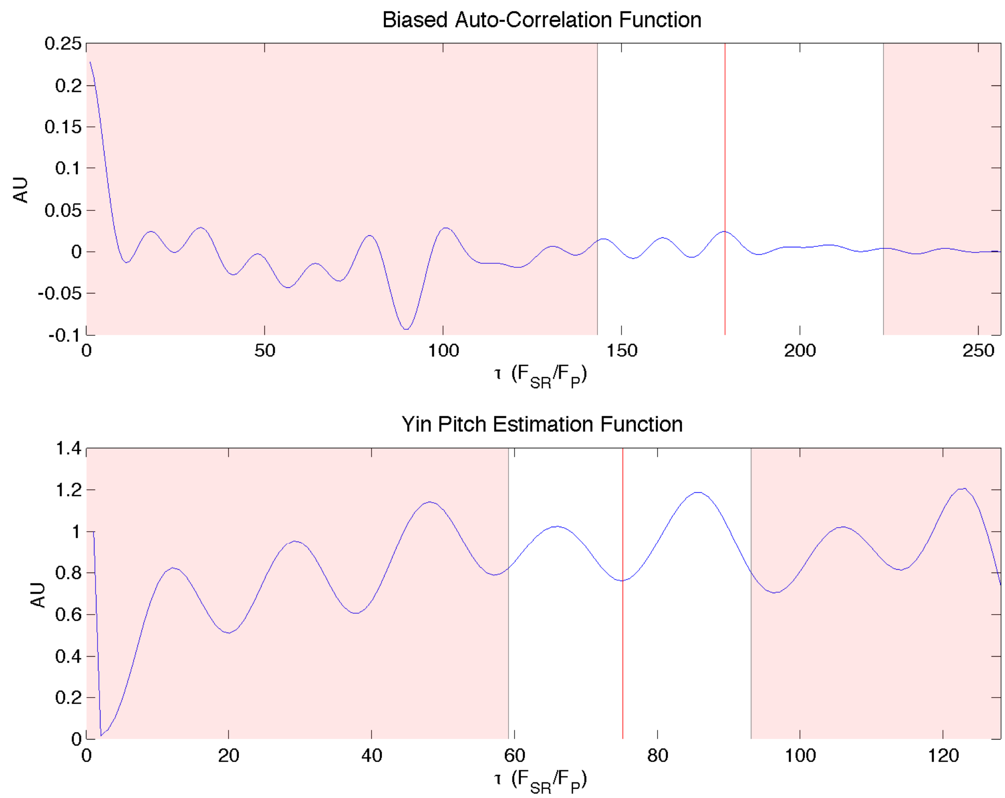


Figure 5.6: Examples of biased autocorrelation (top) and Yin difference function with cumulative mean (bottom). Both examples, taken at different points in time, use $W = 256$ and f_s of 44.1kHz.

for the minimum to be within a given range of the estimate from the fingerboard sensor. The Yin search uses the same search intervals as the assisted autocorrelation. Similarly, we apply the same concept of looking for second harmonics when the expected period is otherwise too long for reliable estimation, and we finish with parabolic interpolation. The use of the difference function, Equation 5.2, means Yin requires twice the window size as biased autocorrelation, effectively doubling its minimum detection frequency. As such, when the sample window is small and the frequency is too low for even a second harmonic Yin estimate, we use the raw sensor pitch estimate.

Example

Figure 5.6 demonstrates the advantages and mechanisms of sensor assistance. While the algorithms would normally search across the whole function, the fingerboard sensor informs us that the pitch must lie in the non-shaded area. In the autocorrelation example (top), the hand labeled target pitch is 246.9Hz, as marked by the red line. The g-string gives us a normalized voltage reading of 0.778 which, using Equation 5.4, provides a hardware pitch estimate of 243.1Hz. The major third either side of the hardware estimate defines the autocorrelation search area of 194Hz-304Hz, corresponding to a τ between 145 to 227. The algorithm then searches that τ range to find the maxima, resulting in a pitch estimate of 247.2Hz.

In the Yin example (Figure 5.6 bottom), the hand labeled target pitch is 293.7Hz. This frequency is actually below Yin's range when using a 256 sample window at 44.1kHz. However, the finger sensor input tells us the expected frequency is 296.6Hz so we define a minima search around the second harmonic, 593.2Hz, or $\tau = 74.34$. We find the minima at $\tau = 75$ which then undergoes quadratic interpolation and is doubled to provide a pitch final pitch estimate of 295.0Hz. Both examples in Figure 5.6 are instances where the pitch estimate would be incorrect without the assistance of fingerboard hardware.

5.5 Testing and Results

5.5.1 Test Setup

All low latency pitch detection tests and results here used the original sensor design with the strings as conductors.

Tests were conducted by recording a performance and the corresponding data feed. In order to directly explore differences in performance for variable window size, results were collected post-performance using the original audio and data for repeat processing. The audio and data samples were synchronized by recording a low sample rate version of the audio as part of the sensor feed and then matching onsets in the low sample rate audio with the regular 44.1kHz sample rate audio. The session was then played back with live analysis of the audio informed by the sensor data measured at the corresponding time during the original recording. The restricted pitch detection method is equally suited to and used in real-time in live performance, though latency from gathering the sensor data over USB should be considered and the audio delayed to compensate if needed.

Three sessions of 2-3 minutes made up of 40, 137, and 202 notes were evaluated. Two of the segments were recorded at 44.1kHz, and one at 48kHz. The segments consisted of scales and arpeggios spanning all notes in G major in first position range of the violin (G3-B5). There was a weighting towards lower notes on the D and G string (under 440Hz). As the performance was to be hand labeled using standard pitches, the violinist was asked to avoid vibrato and to try to minimize holding multiple fingers down on different strings since the current system only presently supports monophonic performance. They were otherwise free to play normally.

Pitch estimates were collected for window sizes of 128, 256, 512, 1024, and 2048 samples. In each case, the hop size was set to one quarter of the window, with the exception of the 2048 sample window which, for on-line computational reasons, had a hop size of one half the window. For $W < 512$, the algorithm used a search window of $\pm$ 12.5% (just intonation whole-tone) around the fingered expected pitch and a $\pm$ 6.7% (just intonation semitone) around expected open strings. Otherwise, the search region was within a perfect fourth.

Results were filtered to exclude periods of silence (defined as less than -48db) and the result from each hop compared against two sets of labels. The first were hand labels of the expected pitch, and the second set was from a 2048 sample window, 64 sample hop-size, Yin-FFT analysis using the Aubio pitch detection plugin by Bossier and Cannam in Sonic Visualiser [22][1].

The reason for the two comparison sets is that the audio is a human performance on a violin. Hand labels matched intended pitch, which may differ from actual pitch. This may be because of performer error or instrument tuning. For instance, in one session, the entire violin was out of tune by 40 cents. As performer pitch error would influence scoring of pitch estimates, a Yin estimate at a high window size was also used, as Yin is widely accepted as a solution to (high-latency) monophonic pitch tracking. The Yin estimate will itself have errors as Yin is not 100% accurate– for instance, note changes typically result in momentary loss of valid estimate and a larger window will be less responsive to small momentary pitch variations– however, when stable, the Yin label is expected to be more accurate than the hand label.

Comparissions were made within 10 (0.57%), 30 (1.75%), 50 (2.93%), and 100 (5.95%) cents of the central pitch. Some of these are quite tight tolerances but were chosen based on psycho-acoustic tolerances. Pitch differences within 10 cents are distinguishable but tolerable and an estimate within 50 cents should round to the correct MIDI pitch, while 100 cents is the nearest equally tempered semitone.

5.5.2 Results

Results for window size, W, of 128, 256, and 512 are given in Table 5.1[2]. These window sizes convert to window lengths of 2.9ms, 5.8ms, 11.6ms at 44.1kHz, and 2.7ms, 5.3ms, 10.7ms at 48kHz respectively. Window lengths of 1024 and 2048 samples, 23.2ms and 46.4ms, were calculated for reference and discussion, but are not considered fast enough for low latency

[1]The version of the Aubio plugins used to generate Yin-labels was subsequently found to have a bug in interpolation which caused error around the correct pitch.

[2]Yin-labeled results for 128, and 256 sample window sizes in Table 5.1 were generated using the Aubio plugin with a bug, meaning ground truth was incorrect. Results for the 512 window were re-run and improved using a corrected version of the plugin.

use.

It is clear that for shorter windows under 512 samples, sensor-assisted pitch estimates outperform traditional “biased” autocorrelation and Yin. Within 100 cents using a 128 sample window, assisted autocorrelation and assisted Yin triple the accuracy of traditional methods. With a 256 sample window, assisted pitch detection still significantly outperforms existing means: assisted Yin offering a 25% improvement, and assisted autocorrelation offering a 48% improvement over standard Yin (100-cent accuracy, Yin labels). Increasing the window size to 512 samples, the advantage of sensor assistance is reduced and accuracy differences become insignificant.

At window sizes above 512 samples, assisted pitch-tracking tends to slightly outperform Yin when evaluated using hand labels: .956 (AC+S) vs .941 (Yin) accuracy at 50 cents, and .952 (AC+S) vs. .928 (Yin) accuracy at 50 cents with a 2048 sample window. This also holds true using Yin-FFT annotations: .939 (AC+S) vs. .931 (Yin) within 50 cents with a 1024 window, and .926 (AC+S) vs .901 (Yin) within 50 cents with a 2048 sample window. It is possible that comparing against Yin annotations will give Yin results an artificial advantage, since both the algorithm under test and the annotations may produce the same errors.

An additional advantage of sensor assisted pitch detection for some contexts is that the error range is much more limited. As visible in Figure 5.7, for fingered notes, error with assisted autocorrelation is limited to within a minor third with the only large error occuring during an open G-string. In contrast, the raw Yin estimate fluxuates dramatically across more frequencies.

5.6 Discussion of Fingerboard Testing

While the 66.8% hit rate using a 128 sample window is too low for practical use, using assisted autocorrelation with a 256 sample window is promising. The advantage the sensor information gives the algorithm over an uninformed estimate can be made clear by considering Figure 5.7. First, the search limits provided by the sensors help screen out noise during a pitch change so that the new pitch is typically found faster. As visible in Figure 5.7, the loss of pitch estimate during transitions is easily visible with Yin, but far less so with assisted autocorrelation.

	128 Sample Window							
	Hand Labeled				Yin Labeled- 2048 buffer			
	Accuracy within # cents				Accuracy within # cents			
Alg.	10	30	50	100	10	30	50	100
AC	.145	.227	.234	.236	.044	.054	.054	.054
AC+S	**.268**	**.483**	**.559**	**.668**	**.331**	**.574**	**.713**	**.798**
Yin	.103	.171	.178	.180	.056	.058	.059	.064
Yin+S	.095	.227	.317	.527	.089	.165	.244	.449
	256 Sample Window							
	Hand Labeled				Yin Labeled- 2048 buffer			
	Accuracy within # cents				Accuracy within # cents			
Alg.	10	30	50	100	10	30	50	100
AC	.502	.505	.511	.511	.252	.260	.260	.260
AC+S	**.537**	**.769**	**.801**	**.819**	**.706**	**.832**	**.859**	**.877**
Yin	.344	.528	.552	.558	.650	.661	.662	.696
Yin+S	.336	.552	.611	.687	.596	.623	.647	.740
	512 Sample Window							
	Hand Labeled				Yin Labeled- 2048 buffer			
	Accuracy within # cents				Accuracy within# cents			
Alg.	10	30	50	100	10	30	50	100
AC	.507	.720	.736	.741	.441	.711	.736	.39
AC+S	**.660**	**.930**	**.954**	**.962**	.618	.914	.941	.946
Yin	.656	**.930**	.952	.959	**.632**	**.920**	**.948**	**.952**
Yin+S	.649	.923	.949	.957	.629	.908	.937	.940

Table 5.1: Pitch detection accuracy for 3 window sizes of 128, 256, and 512 samples. Comparisons use both hand-labeled performance data and a 2048 sample Yin-FFT pitch detection analysis.

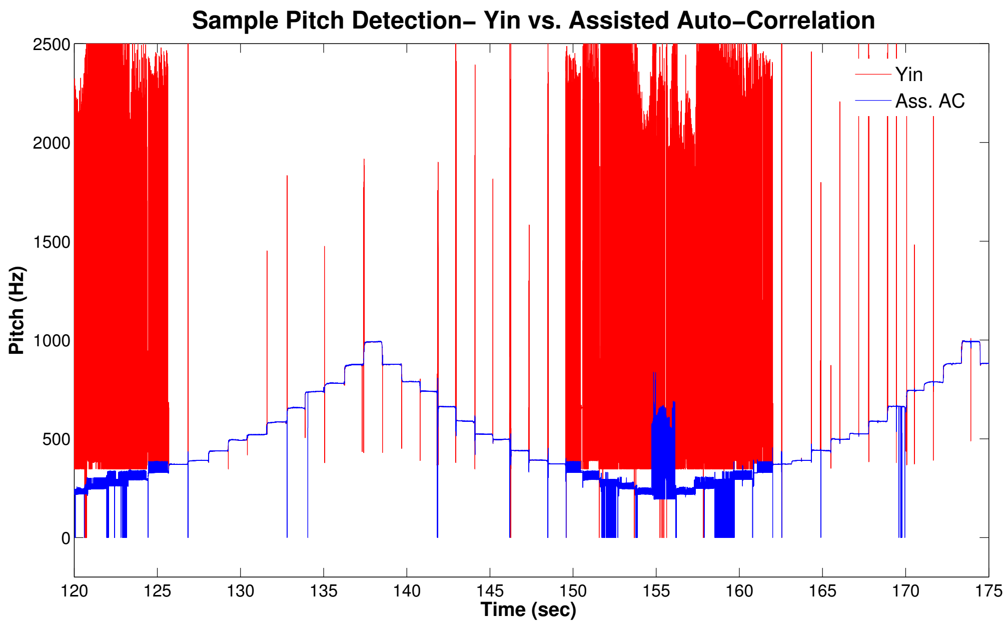


Figure 5.7: Pitch estimates derived using Yin and assisted autocorrelation with a 256 sample window and major third search area at 44.1kHz for an audio segment consisting of scales. The sensor assisted results have significantly less variation in error.

This additional improvement in accuracy when finding a new pitch directly enhances the low latency performance.

Next, a 256 sample window at 44.1kHz corresponds to a 5.8 ms window so that no frequency under 172 Hz will complete a wavelength within the window. Additionally performance will improve with multiple wavelengths. Yin has a minimum detection period of $\frac{1}{2}\tau_{max}$ where τ_{max} is limited to the length of the window. Hence, with a 5.8ms window, Yin will fail to detect frequencies under 344Hz, a result clearly illustrated by Figure 5.7. However, sensor assisted autocorrelation is able to reasonably estimate the correct pitch, plus it is never further off than a minor third and rarely significantly underestimates the frequency.

The issue with the original sensor arrangement in Section 5.3.2 using conductive strings, where a fingered note did not always produce an electrical connection, was identified during this test and found to be particularly problematic on the thin E-string, where the string can make a groove in the finger and not contact the fingerboard. An example of failed electrical contact affecting pitch estimates can be seen on the right side of Figure 5.7, at around 170 seconds. Identifying and removing instances where the string unexpectedly lost electrical contact with the fingerboard improves estimation accuracy roughly 1.5-5.0% over present accuracy for both modes of assisted prediction. This improvement is enough that if the contact error can be eliminated, assisted pitch detection was expected to outperform Yin for all window sizes and was the motivation for switching to the second sensor design.

Beyond refining sensing arrangements, a major opportunity for improving results is to fully calibrate the hardware sensor for correct pitch. At low window sizes, accurate pitch estimate turned out to be heavily reliant on hardware estimate accuracy, but the initial tests did not fully exploit hardware sensor accuracy and stability. During formal testing, we used open string voltage as v_{MS} in Equation 5.3 and did not diligently tune the hardware beyond ensuring that fingering near the nut would be minimally detectable. Of the three test sets analyzed, only one had an average hardware pitch estimate error within 100 cents of labeled pitch, thus enabling us to experiment with a semitone search window. With a 256 window size and the semitone search window, this set was 91% accurate within 100 cents using assisted autocorrelation and 89% accurate using assisted Yin. Three to four% of the error was due to contact error, with much of the remaining error occuring during open G-strings. Presumably, if we had calibrated hardware estimates to be closer to expected pitch, we would have documented higher accuracy at low latencies.

An additional hardware challenge is that velostat is temperature sensitive so its resistance is not stable. This was dealt with by adding a potentiometer at R_c in Figure 5.4 to control the current supply enabling us to vary the drive current to calibrate the fingerboard sensor. We also found that with the first sensor design charge appeared to build in the velostat when left on for a long time, possibly due to interaction with the wire wrapping of the conductive strings acting as inductors. This further altered conductivity but could be countered by turning off power for a while.

The issues with missed contacts and held charge were eliminated or at least significantly reduced by switching to the second design, Section 5.3.2. Additionally calibration was dramatically improved. Subsequent calibration was done by making adjustments to Equation 5.4 for expected areas of error such as additional velostat length between the bottom of the conductive trace and the electrical connection of the velostat for each physical sensor build. Prior to every use Figure 5.4's R_c was expected to be adjusted so that hardware estimates for fingerings of specific pitches matched within ± 15 cents.

5.7 Adaptation of Restricted Pitch Search for Real-World Environments

All of the pitch detection work presented so far has assumed exclusively monophonic performance. This allowed for simplicity in the test environment; for instance, the algorithm for deciding finger placement merely picked the highest fingered position. However, even excluding intentionally polyphonic violin performance written to utilize chords, real-world violin playing is rarely clean enough to treat as perfectly monophonic. It is not only common to unintentionally brush other strings while playing and for the left hand to inadvertently make contact on non-played strings, but it is also frequently intentional and correct to leave fingers down on unplayed strings. Additionally, looking at the results of low latency pitch detection, pitch estimation struggled to correctly identify which open string was being played: an area of necessary improvement.

Before trying to use low latency pitch estimation with string players, we wanted to improve robustness to polyphonic inputs and real world fingerings. Two implementations were designed, the first, pictured in Figure 5.8, based on the electric violin but using a polyphonic bridge. The polyphonic bridge allowed largely independent pitch analysis of each string. Second, an acoustic violin with more advanced software for correctly recognising pitch with relaxed performance requirements was built. Both these designs used the second fingerboard design based on the custom built linear potentiometer and the same algorithms for hardware estimate. They only differed in the violin and the audio analysis. While we did not carry out any formal performance tests on either, the polyphonic version was used in Chapter 8 and the acoustic version was used in Chapter 9. Here we include an informal assessment.

Figure 5.8: Electric augmented violin with polyphonic input and augmented bow.

Due to the potential for high computational loads, real-time performance versions of the pitch detection algorithm presented in Section 5.4 switched from pitch correlation using raw multiplication to correlation based on the signal's fast fourier transform (FFT) which is faster ($O(nlogn)$ vs $O(n^2)$):

$$Corr(r_t, r_t) = R(\omega)R^*(\omega) \tag{5.6}$$

where $R(\omega)$ is the fourier transform of r_t and $R^*(\omega)$ is the complex conjugate of $R(\omega)$. Both methods yield equivalent results. For more information on logistical software implementation of real-time systems, please see Section 7.1.

5.7.1 Improved Pitch Detection Using a Polyphonic Bridge

A useful way to address polyphonic inputs was to swap the electric violin's regular bridge with a polyphonic bridge. The polyphonic bridge allows us to separate the audio inputs of each string so each string can be given its own audio channel. Though there will be crosstalk from other strings, the dominant signal on a single string's audio channel is its designated string.

The Barbera Polyphonic Bridge

The polyphonic bridge is a passive multi-transducer violin bridge made by Barbera Transducer Systems[3]. We are using a solid body twin hybrid model bridge which has a separate, dual piezo transducer elements for each string providing a single audio output for each string. The solid body is reportedly more resistant to cross-talk, though we found still significant albeit acceptable cross-talk. Barbera's bridges are well respected within the electric violin industry, used by Zeta Midi Violins, Wood Electric Violins, Vector Electric Violins and others. The bridge is paired with an EMG 6-string polyphonic pre-amp which provides individual and summed output.

Sensor Assisted Pitch Estimation Using Polyphonic Input

Low latency sensor assisted pitch estimation using a polyphonic input uses the same core algorithm was the one presented and tested in Section 5.4 but with fingering inputs from each string, paired with the audio input from each string.

[3] `http://www.barberatransducers.com/violinpickups.html`

Returning to the fingerboard sensor in Section 5.3.2, the fingerboard sensor is designed to capture the bridge-most finger on the fingerboard for each string as, apart from techniques using partial fingerings like harmonics which we do not detect, that placement will determine the fingered string's resonating pitch. This string specific fingering is paired with the string specific audio supplied by the polyphonic bridge to produce a string specific pitch estimate. Each string having its own audio signal, we must calculate four complete pitch estimates. We then use comparative string volume to select the primary performed pitch.

This is accomplished by using sensed finger placements to calculate a hardware pitch estimate, as in Section 5.3. Previously, we looked through the separate fingerboard inputs and chose the one performed on the highest string and assumed an open string if there was no fingered input. With the polyphonic bridge, we use all fingerboard sensor inputs to produce four separate hardware pitch estimates. Again, if there is no finger detected, we assume the open string, but this time, we only search the given interval around the specific unfingered string. The polyphonic input estimation algorithm uses the biased autocorrelation for the core pitch detection algorithm.

For volume we use the signal root mean square (RMS) calculated either directly or using the FFT, $\hat{R}(\omega)$. Once we have the pitch estimates for each string we treat the target audio as if monophonic. We use the RMS volume to identify which string was loudest assume it is the played string, and use the pitch estimate from that string as the played pitch. With the polyphonic bridge enabling an independent pitch estimate for each string, it would be presumably simple to do a polyphonic treatment of the performance using all four pitch estimates and their RMS.

$$RMS = \sqrt{\frac{\sum_{t=1}^{N} r(t)^2}{N}} \qquad \text{or} \qquad RMS = \frac{\sqrt{\sum_{\omega=1}^{N} |\hat{R}(\omega)|^2}}{N} \tag{5.7}$$

Informal Assessment

Informal experiments with the polyphonic violin demonstrated success at eliminating instability in the pitch estimate on open strings. In instances where the player brushes another string creating non-melodic audio, the primary line is still correctly identified and followed by the pitch estimate. Additionally, it correctly handled fingers placed but not played. Test samples showed pitch was correctly estimated across the full tested sample range (MIDI notes G3 - D6) within or above the expectations of earlier tests. As it is inherent in the design, polyphonic input assisted pitch estimation will suffer when finger contact with the fingerboard is insufficient.

The sound produced by the polyphonic bridge itself is a bit harsh and can sound very thin. Presumably, matched to a more resonant body and with added signal effects, it can be enhanced to sound better, but work on improving pick-up sound quality was beyond the scope of this thesis.

See Chapter 8 for user reaction to playing on the polyphonic bridge augmented violin.

5.7.2 Monophonic Pitch Detection for an Acoustic Violin

While tests earlier in this chapter and the polyphonic version use an electric violin, a simple fact is that acoustic violins are far more common. For most people, the acoustic violin is what they own, the sound that is familiar, and the instrument they prefer. As a goal is real-world useability, adapting the restricted pitch search methods for an acoustic violin greatly widen applicability.

Selecting String Based on Estimate Confidence

Again, the fingerboard in Section 5.3.2 was designed to work on any fingerboard, electric or acoustic, so we can continue to use Equations 5.3 and 5.4 to calculate $\hat{f}_{hw}$. As with the polyphonic bridge implementation in Section 5.7.1, we treat each string separately, but this time using the violin's acoustic audio. For the audio input any standard violin microphone should work; we use a electret microphone designed by Mark Feldmeier[4] mounted using the strings below the bridge as in Figure 5.9. A single autocorrelation of the audio is executed and then a restricted search is performed on the autocorrelation results for each $\hat{f}_{hw}$ derived from the finger closest to the bridge for each string individually. Again, if there is no finger

[4] `http://www.openmusiclabs.com/`

on a string, we assume the string must be open. Without the ability to calculate an RMS for each string individually, we must instead rely on pitch estimation confidence.

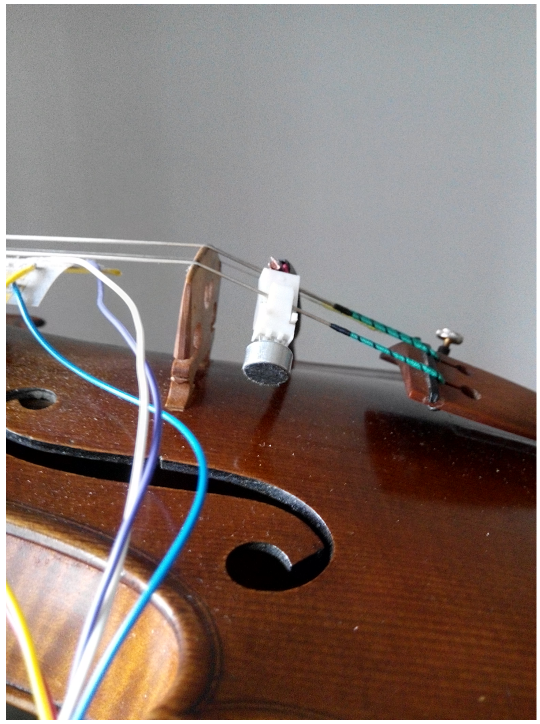


Figure 5.9: Custom electret microphone clipped to string providing monophonic audio input for an acoustic augmented violin.

With τ_θ the non-interpolated version of the pitch estimate selected according to Section 5.4, the pitch estimate confidence, ϕ, is measured as $\phi_t = r_t(\tau_\theta)$ (Equation 5.1) when using restricted biased autocorrelation or $\phi_t = 1 - d'_t(\tau_\theta)$ (Equation 5.5) when using restricted Yin. As mentioned in Section 5.6, when properly calibrated the fingerboard is typically accurate within ± 16 cents. An uncertainty of ± 16 cents allows us a much narrower search band around the hardware estimate, enabling the acoustic violin version of low latency pitch detection to only search within a whole tone either side of $\hat{f}_{hw}$. We compare the estimate confidences for the restricted search on each string and pick the one with highest confidence, ϕ_{max}. With ϕ_s the confidence for each string:

$$f_{est}(t) = \arg\max_{\phi_s} \left(\phi_s, \hat{f}_s(t)\right) \quad \text{where } s = 1...4 \tag{5.8}$$

Unlike pitch detection using the polyphonic bridge, use of the acoustic audio only is less likely to cope with correctly handling polyphonic audio, but the improved pitch detection methodology does have improved resistance to accidental polyphony. We have not experimented with using it to detect polyphonic audio.

Informal Assessment

Again, no formal testing was performed to verify the accuracy of restricted pitch estimation on the acoustic violin. Open strings were detected correctly although the current algorithm sometimes struggled with the bottom of the G-string, 196 Hz G and 220 Hz A (MIDI notes G3-A4). The algorithm also demonstrated resilience to fingers placed but not played. However, using restricted biased autocorrelation, the algorithm struggled to detect specific frequencies. For instance, at a 48KHz audio sample rate, the 440 Hz A5 suffered instability, which stabilized after switching to 44.1kHz, but then the 330Hz E4 estimate became unstable. Switching to restricted Yin reduced overall accuracy on the G-string (MIDI notes G3-B4), but eliminated instabilities seen in higher frequencies. Otherwise, both pitch estimation methods gave correct estimates within the expectations of our design. As such, we are currently using restricted Yin as simple songs played by beginners are often centred around the A string as error above the D-string for our main test audience would be more distracting than additional errors on the G-string.

One difference in behavior between the electrical and acoustic estimates is error response. With the electric violin, if finger contact is lacking, the pitch estimate jumps constantly so that pitch error always results in instability. However, on the acoustic violin, when contact with the first finger is lost, a stable incorrect estimate is given at the boundary of the semitone search area. To combat this particular, potentially confusing error, the search area above the open string is increased to a minor third ensuring a major second above the string is included in the restricted search. The increased search area means that there is a higher risk of inconsistent error at lower latency, but a lower risk of a consistent error.

Further assessment and user reaction to using the real-time acoustic augmented violin can be found in Chapter 9.

5.8 Conclusions

Combining audio analysis with sensor input allows us to significantly improve on existing pitch tracking methods in low latency contexts. We illustrated two methods for designing a velostat based linear potentiometer for the fingerboard along with supporting electronics. Our custom built fingerboard sensor demonstrated excellent rough accuracy when estimating violin pitch, producing reliable estimates within 20 cents of the correct pitch. Our sensor designs are non-intrusive but pitch estimates are reliant on correct tuning of the violin.

In isolation, audio pitch tracking algorithms are precise, but suffer from occasional large harmonic errors and struggle to perform reliably at low latencies. Because the errors are of different types, pairing the fingerboard sensor with audio based pitch estimation algorithms, uses the strengths of both methods to accommodate for the weakness of the other.

We were able to improve from a nominal 5% detection of correct pitch within 30 cents when using the popular Yin [34] algorithm (128 sample, 2.9ms window) to 57% detection rate using sensor informed autocorrelation with the same sample window. Table 5.1 also demonstrates that with a 256 sample 5.8ms window, detection of correct pitch within 30 cents benefits from a 34% improvement using augmented methods instead of Yin. Additionally, for any fingered note, all errors using augmented techniques were entirely within a whole tone of the correct pitch, eliminating octave errors that commonly occur with audio-only analysis [133].

Further, the combination of sensor data and audio analysis allows frequencies to be found that are below the theoretical minimum for a given window size. The hardware estimate is used to identify when the expected frequency is too low and will then search the autocorrelation results for the second harmonic instead of the fundamental. While raw audio analysis was essentially un-useable using a 128 sample window, autocorrelation informed by hardware estimates was still able to correctly estimate within 50 cents of the pitch over 70% of the test set.

We used the combined sensor and audio analysis approach to design two low latency pitch tracking augmented violins. We used an electric violin with a polyphonic bridge to develop the electronic augmented violin, used in tests in this chapter. The electric augmented violin combines the audio from each string with the finger placement nearest the bridge for that same string to estimate the pitch for the string. The loudest string is chosen for the estimated pitch. We perform a full user study in Chapter 8. The acoustic augmented violin, uses a single monophonic input which is paired with the finger placements for each string to generate pitch estimates for all four strings. The estimate with the highest confidence is selected as the estimated pitch. Informal testing of the acoustic electric violin found its performance was equivalent to the electric augmented violin. These two low latency pitch estimate violins enabled us to perform studies of real-time pitch correction in Chapter 8 and Chapter 9

Chapter 6

Case Study: Note Onset Using Sensors

Section 6 incorporates significant material from the article 'A Low-Cost Real-Time Tracking System for Violin' by Pardue, Harte, and McPherson originally published in JNMR 44.4 [134].

This chapter presents a case study on note onset detection. In order to demonstrate and test the augmented violin, we tackled a traditional music information retrieval problem of note onset.

Note onset detection is useful for a variety of tasks such as automated transcription [6], score-following [29], and performance analysis [61]. Though audio-based note onset is fairly mature for percussive instruments, it remains an open challenge for many non-percussive instruments [13]. Real-time note onset detection provides a useful demonstration of our real-time violin tracking methods since it requires recognizing note changes brought about by actions using the bow and/or fingerboard.

6.1 Non-Percussive Note Onset Detection

For instruments with clear attack transients like piano and percussion, detecting note onsets from audio is considered largely solved [27]. However the magnitude-based methods used for percussive instruments do not work as well for instruments with slow or subtle onsets such as voice, wind instruments and bowed string instruments. [38] and [27] provide overviews of different approaches to onset detection.

Methods for non-percussive note onset detection commonly focus on differences in the spectral energy or frequency difference between windows [28], phase differences that might suggest a new waveform [6], using the mel band to evaluate differences in spectral content as a more psycho-acoustically relevant technique [87], and more complex means that blend multiple techniques to improve results. Two problems for non-percussive note onset detection are latency and accuracy. Many onset algorithms are non-causal, though recurrent neural networks have showed promise for real-time detection [12].

Attempts to find non-percussive onset in real-time typically use spectral analysis [27]. Though they are real-time, latency is high. Not only are the minimal sample windows required for computing accurate spectral content typically over 10ms, but the need to compare changes across successive windows will also further push latency well above the target of 10ms for real-time performance systems [49]. A recent comparison of real-time onset detectors requires that onsets are detected within ±25ms of their ground truth labels, but it uses windows of 46ms (2048 samples) with 10ms between hops [13], implying a minimum average latency of 33ms.

Accuracy also remains below usable standards for controlling sounds in a live performance. Spectral methods can be confused by vibrato, trills and other expressive devices. Slow pitch changes can lead to both missed onsets and false positives, while vibrato typically leads to a significant increase in error. *OnsetDetector* from [164] is one of the better tools for bowed onset in monophonic pitched contexts, achieving around 70% F-measure (defined as $F = \frac{2PR}{P+R}$ where P is the precision, the percentage of true positives, and R is the recall, the percentage of positives found). Böck's *SuperFlux* algorithm was significantly improved by adding vibrato suppression [11]. With a 25ms window, it achieved an F-measure of around 75% with 83%

precision and 69% recall.

6.1.1 Violin Onset Detection using Sensors

Although we have not seen work focused on using inexpensive[1] sensors with a violin specifically for detection of note onset, there are other approaches that may be able to yield similar or complimentary results. If working well, any of the fingerboard sensing techniques mentioned in Section 3.1.1 would be theoretically capable of capturing similar fingerboard data to that used within this case study.

For onsets detectable through bow motion, it is possible to obtain motion data by using gyroscopes for tracking basic forearm motion [74], or mounting an accelerometer on the frog [185]. Both these techniques provide only limited information about bow position but have been used effectively for tracking bow changes and do not require placing sensors on the violin. Adding ultrasound sensors [125] or IR lights and an IR camera [110] to both the violin and the bow, or bow hand have also been used for tracking bow position in relation to the violin. However a major constraint on these approaches is that none of them are capable of tracking actual string contact or bow pressure.

6.2 Mechanics of Violin Note Onset

On the violin, two things are required for a bowed note to sound: bow velocity and downward bow force on the string. New notes come mainly from change of bow direction, change of string (correlated with bow angle with respect to the violin), or a change in left hand finger placement. Sensors can capture many of these actions, particularly changes in bow velocity and pressure that subsequently lead to sound being produced [36], allowing early identification of note onsets. String changes may be challenging to identify quickly from bow position and pressure alone, but audio analysis may be able to fill in this gap. Audio analysis may also help identify false positives or even missed detections in sensor data.

[1] We are excluding methods using EMF and specially designed motion capture systems due to their prohibitively high price.

In the following case study, we consider five types of note onset:

1. Off-string attack (OSA): bow is placed on the string and moves.
2. Bow change (BC): bow is already in contact with the string but the bow changes direction.
3. Finger change (FC): pitch change of at least 1 semitone through left-hand finger change.
4. String change (SC): pitch change through change of string the bow is playing.
5. Slurred repetition (SR): delineation of new notes through accent accomplished by abrupt change in bow pressure and/or velocity.

While only one of these actions is necessary to define a new note, they often occur in conjunction. Identifying these actions requires fingerboard tracking and bow tracking. Here, we consider four of the five note onset cases separately (excluding string changes) and assess the ability of the augmented violin to detect note onset in real time. Informal tests suggest that string change can be detected with a gyroscope on the frog of the bow or a bridge which provides separate pickups for each string, but this is beyond the scope of this study.

6.2.1 Off-string Attack

This category of note onsets includes any situation in which the bow first makes contact with the string when having previously been in the air. It includes setting the bow firmly on the string and then moving it (typical for accents), initial contact with string while the bow is already in motion, and bouncing the bow on the string (*spiccato*).

Optical sensors on the bow easily detect whether the bow is on the string, and thus can detect whether an off-string attack may be happening. Contact with the string causes the average hair-stick distance measurement across the four sensors to deviate from the off-the-string measurement (which is also used to estimate bow tension). Placing the bow on the string will typically result in at least a 10% increase in the instantaneous average while all but the most aggressive waving of the bow in the air will result in only a 3-4% deviation in the measurement of the raw tension.

We use the pressure estimate from Chapter 4 to detect bow-string contact; this estimate is less dependent on the location of the contact along the bow than raw sensor readings. We chose 60g (0.59N), corresponding to 15% above off-string sensor readings and slightly less than the weight of the bow, as the threshold for bow on the string; due to hysteresis, we chose 45g (0.44N) as the threshold for the bow coming off the string.

To detect bow motion, we simply take the difference between successive bow locations. This imposes an implicit latency of 4ms in detecting off-string attacks. In order to avoid false triggers around the same event, no off-string note onset is considered within 100ms of the previous onset. This corresponds to a realistic maximum rate of note production for almost all performance scenarios.

6.2.2 Bow Change

Bow changes are variations in direction of movement while the bow remains on the string. These appear as local maxima and minima of bow position over time (see Figure 6.5 in results). Finding these extrema in real time is more difficult; extrema can only be identified retrospectively, and noise in the position estimates imposes a tradeoff between latency and accuracy depending on the number of successive frames examined.

Bow changes can only be detected when the bow is known to be on the string, i.e. when an off-string attack has previously been detected. Upon detection of an off-string attack, the initial stroke is identified as up-bow or down-bow. Changing from up-bow to down-bow implies a local minimum in position; changing from down to up implies a local maximum.

Figure 6.1 shows the bow change detection procedure. Position data is filtered to remove high-frequency noise, and filtered bow position is then examined within a sliding historical window of 15 samples. To find a down-to-up change, we look for the first instance where the most recent position sample is below the earliest sample in the window. The bow change is then identified as the location of the local maximum in that window. A similar procedure is used for up-to-down change. To reduce false positives due to the position estimate getting stuck (Section 4.5.3), the window must include a clear change in bow position defined by motion passing a minimum threshold. The start of the new bow stroke must also be in a

{Main Loop for Down-bow} ▷ For a down-bow we look for local maxima.

if $currentBowPos < bowPosHist[lastInWindow]$ **then**

  $trend \leftarrow trend + 1$

  if $((trend > 5) \quad and \quad ((max(bosPosHist) - min(bosPosHist)) > minThresh))$ **then** ▷ We declare a hit!

    $bowChange \leftarrow argmax(bosPosHist)$ ▷ Declare change at the window max.

    $detectionTime \leftarrow now$ ▷ When are we labeling the bow change?

    $bowDir \leftarrow upBow$ ▷ Switch to Up-bow Loop and look for minima.

  end if

else

  $trend \leftarrow 0$

end if

$bowPosHist \leftarrow bowPosHist.push(currentBowPos)$

$bowPosHist \leftarrow bowPosHist.pop(lastInWindow)$

Figure 6.1: Pseudo-code describing the algorithm for finding the transition from a down-bow to an up-bow. The transition from up to down-bow looks for a minimum instead of maximum.

physically plausible location compared to the start of the previous bow: for example, when changing from up-bow to down-bow, the down-bow must start at a position closer to the frog than where the up-bow started.

A shorter time window will reduce latency of identification at the cost of greater noise susceptibility. We also reduce noise susceptibility by requiring several successive samples to confirm the direction change, though this adds further latency. We chose to require five successive confirmations of the direction change. For symmetrical alternating bow strokes, the extremum will fall roughly in the middle of the window, introducing $\sim$ 50ms of total expected latency between the physical bow change and labelling. The latency is offset by the fact that the bow change will precede the audio onset since the string must be re-excited. Bow change latency is thus still above our targets, but further reductions in sensor noise will allow tighter windows and lower latency. No two bow changes can be detected within 125ms of each other.

6.2.3 Finger Changes

Note onset from pitch change is caused by either string changes (not considered here) or change in left-hand finger placement. To detect the latter, we use hardware pitch estimates from the fingerboard sensor (Chapter 5.1). Position readings from the finger closest to the bridge are converted to a frequency estimate using Equation 5.4. For onset detection, pitch accuracy is less important than detecting changes, so we do not need the assisted autocorrelation technique.

The frequency estimate is linearized by converting it to the number of fractional semitones. It is then high-pass filtered for frequencies above 1Hz to emphasize instantaneous transitions. Because the resulting filtered equation may decay but not fully return to zero between onsets, we take a high order differential comparing with 40ms previously and looking for transitions that pass a minimum threshold of $\frac{2}{3}$s the instantaneous change expected from a semitone. The lower threshold allows transitions which do not change instantaneously between successive samples. If the sign of the filtered signal matches the sign of the transition, it is declared an onset event. This last restriction avoids false positives due to noise in the decay of the signal after filtering. No new pitch onset is considered for 100ms after the preceding one. Figure 6.2 provides a sample of data used for finger change detection, both raw and processed.

Using the high pass filter, false positives due to vibrato will be eliminated as the pitch change is not fast enough to pass through the filter, nor is the pitch variation significant enough to exceed the threshold. On the other hand, note onsets from slow, continuous glissandi are unlikely to be detected, and we did not design for handling trills. Both glissandi and trills should be detectable with the sensor data, but this is left for future research.

6.2.4 Slurred Repetition

Slurred repetition is when a note is repeated without changing the bow. It is typically accomplished by momentarily reducing pressure and slowing the bow before quickly resuming a higher speed and downward force. It is possible to repeat a note with only pressure or speed, but more commonly the two change together. Focusing only on slurred repetitions which include a pressure change, we can use a variant on the off-string detection algorithm altering

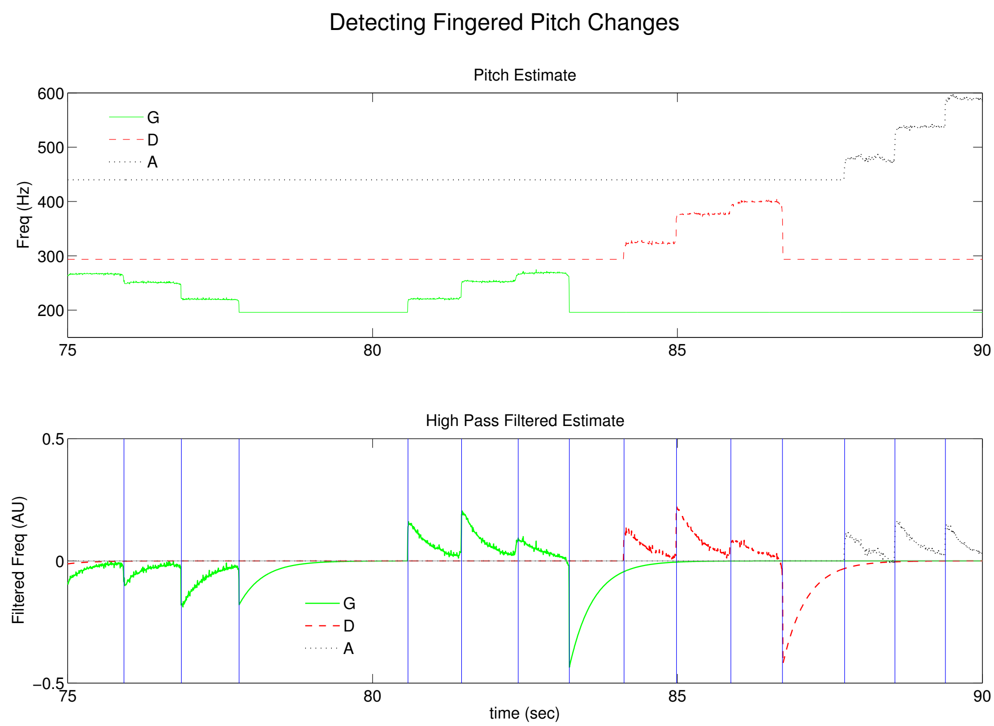


Figure 6.2: Hardware data is used to determine pitch estimates for each string (top) and then run through a high pass filter to emphasize pitch changes (bot). We label a pitch-based note onset (blue vertical lines) when the function exceeds $\frac{2}{3}$ the expected difference due to a half step above/below the prior value.

the thresholds to identify the substantial momentary drop in pressure.

6.3 Testing

Onset detection was tested using three performance cases. The first consisted of three 2-octave G major scales, each with a different bowing style (Figure 6.3): The first scale was played *legato* with two notes to the bow; the second also contained two notes per bow but repeated each pitch across the bow change; the third was played with separate bows, with an off-string *spiccato* stroke on the ascending scale and a more on-string *staccato* or *martelé* on the descending scale. The scales cover each onset case except slurred repetition.

The second case comes from the entirety of the Schubert *Lullaby* from Book 4 of the Suzuki Violin Method (2009 revised edition), a version of which can be found in Appendix A. This piece included slower material and longer slurs. The third case was the first 20 bars of the Seitz *Student Violin Concerto No. 5*, op. 22, also found in Appendix A. It includes a *spiccato* section, a variety of on-the-string transitions including slurred repetition, and faster inter-onset intervals, with notes occurring less than 150ms apart several times. Each of the three cases used a different bow tension, which was measured prior to the selection's start. The total number of different types of onset is given in Table 6.1.

	Type of Note Onset					
Song	OSA	BC	FC	SC	SR	Total Notes
Scales	61	59	66	18	0	163
Schubert Lullaby	4	28	41	7	0	49
Seitz Concerto	46	21	82	24	2	72
Combined	111	108	189	49	2	275

Table 6.1: Number of different kinds of note onsets (on-string attack, bow change, finger change, string change, and string re-attack) in the three sample pieces. Because multiple onset actions can coincide, the sum of the number of onsets per type does not match the total number of notes.

The results presented below were not calculated in real time, but the algorithms were all explicitly written to be real-time capable and causal. Audio from the violin was recorded at 44.1kHz directly to the computer and sensor data was logged. Audio and sensor data were synchronized by capturing a low sample rate version of the audio on the embedded sensor ADC, and then comparing audio markers at the beginning and end of each performance. We found that, based on relative marker locations, audio and sensor streams remained within one sensor sample (4ms) of one another when run in real time. Ground truth note onsets were manually labeled using a combination of visual assessment of audio signal, spectral analysis, and auditory judgment in ambiguous cases.

6.4 Results

Results were calculated separately for each note onset type. They were classified as clear[2] false positives (FP), clear false negatives (FN), and whether they are identified within 75, 50, 25 or 10ms of the labeled onset time or precede the label by more than 25 or 50ms.

Anything below 10ms matches our real-time target while anything within 25ms is more comparable to standard note onset metrics. As we are often detecting the physical actions that lead to sound production, we would frequently expect results to precede the labeled note onset.

6.4.1 Note Onset through Off-String Attack

The effectiveness of note onset detection in the case of an off-string attack is shown in Table 6.2 and illustrated in Figure 6.4. A positive result of the off-string note onset detection was that there were only four clear false positives and one false negative. 86% of onsets are detected within 25ms and 68% within 10ms. 24% of onsets were found more than 25ms in advance, possibly due to the length of time it takes for the string to vibrate and the violin to resonate. Some particularly early detections during spiccato and staccato strokes were likely due to the

[2] Because our real-world test data was sometimes noisy, there were note ons where determining a specific onset point was impossible and similarly, there were playing imperfections that could potentially be interpreted as a note on. In both cases hand labeling of note onset was ambiguous and related classification events were ignored. Only onsets that could be clearly hand-labeled were used.

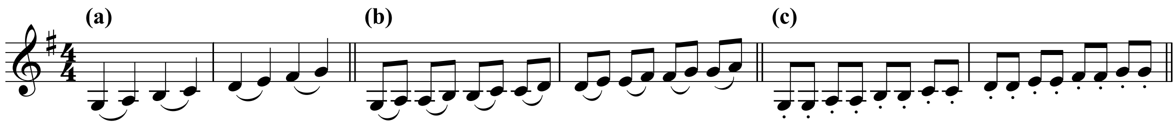


Figure 6.3: Bowing styles used for testing: (a) slurred legato, (b) legato with repeated notes, (c) spiccato. Two octave G-major scales were used in the testing, of which the lower octave is shown here.

			Off-String Attack						Labeled
Song	FP	FN	-50ms	-25ms	10ms	25ms	50ms	75ms	Onsets
Scales	4	1	16	26	43	50	60	60	61
Schubert Lullaby	0	0	0	1	2	3	3	3	4
Seitz Concerto	0	0	0	1	31	42	44	46	46
Combined	4	1	20	27	76	95	107	109	111

Table 6.2: Number of false positive FP and false negative FN off-string attacks along with correct detections found within Xms of the hand labeled note onset. A negative time means the attack was detected prior to the label.

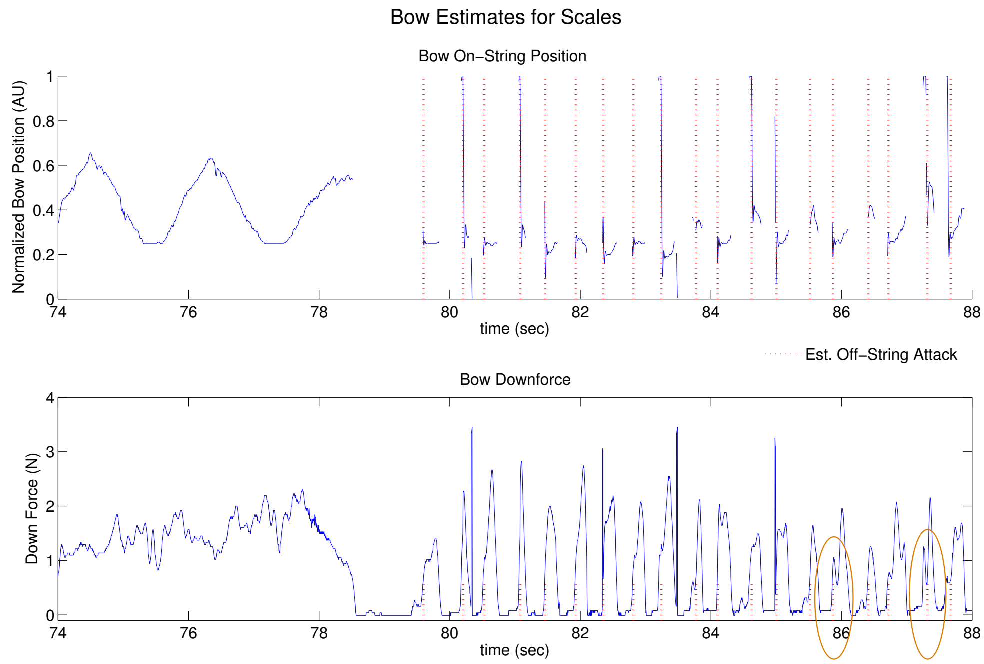


Figure 6.4: Off-string attack classification including spiccato strokes. Force estimates must be above 0.59N (60g) following a period of the bow being off the string and bow movement must also be visible. Small up and down spiccato strokes following note onset are visible. Momentary spikes in pressure are due to estimate error at low applied force. Evidence of grabbing the string can be seen circled later in the sample at 85.87s and 87.31s.

bow being set lightly on the string in advance of the stroke in order to 'grab' it and then holding the bow in place prior to pulling on the string.

6.4.2 Note Onset through Bow Change

Song	FP	FN	**Bow Change** -50ms	-25ms	10ms	25ms	50ms	75ms	Labeled Onsets
Scales	26	0	4	4	6	10	19	29	53
Schubert Lullaby	22	0	4	4	5	7	16	19	29
Seitz Concerto	25	2	6	6	11	13	17	23	35
Combined	73	2	13	14	22	30	52	71	117

Table 6.3: Number of false positive and false negative detected bow changes and correctly detected bow changes found within Xms of the hand labeled note onset. A negative time means the attack was detected prior to the label.

Results for note onset detection through bow change are given in Table 6.3. Due to the impact of noise on real-time recognition, detecting bow change in real time is poor. Although bow direction changes are easy to recognize over longer periods of time, when looking for extrema using relatively small windows in time, false positives frequently occur. If the bow position has any significant instability, it is incorrectly recognized as a hit. Worse, as the present algorithm expects the bow to alternate directions, false positives always come in pairs.

As expected, the latency required to correctly find bow changes was high. Figure 6.5 illustrates classification of bow changes. Altering the detection algorithm to reduce false positives comes at the cost of higher onset detection latency. The effect of the inherent latency due to the real-time constraint is apparent in that only 26% were detected by 25ms after the hand label. Further, removing the real-time constraint and expanding both the search window for extrema to 400ms and the thresholds to declare an event, it is possible to rival audio note onset techniques; for the scale test sample, we eliminate all false positives and false negatives with 89% of bow changes detected no later than 10ms after the hand label. These bow measurements are clearly useful, however improvements in estimate noise will have to be made in order to use optical bow tracking for real-time bow change detection.

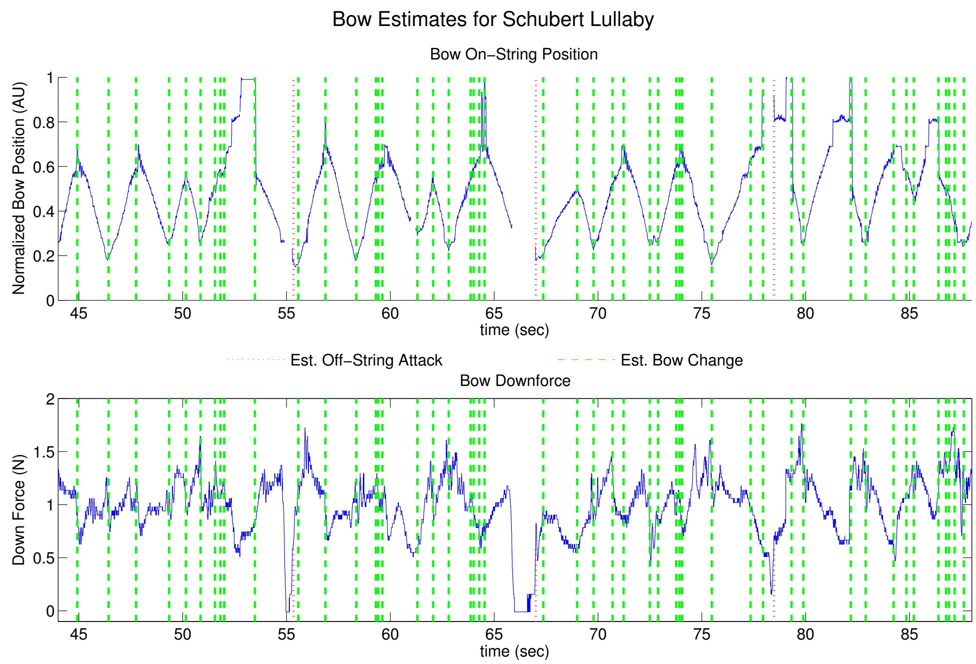


Figure 6.5: Bow based note onset classifications for complete Schubert's Lullaby. Estimated bow changes are marked where the the algorithm places the change, not when it is detected. The position estimate gets 'stuck' around 53s, 79s, 81s, and 86s.

For real-time detection of bow changes, measuring inertial movement through inertial sensors (IMU) may be a better option. As mentioned in Section 3.1.1, the use of an IMU for classifying bow strokes is well established [144, 187] and has in fact been incorporated into our violin tracking hardware, but it is not included in this document as we focus on an optical only approach here.

6.4.3 Note Onset through Finger Change

	Finger Change								Labeled
Song	FP	FN	-50ms	-25ms	10ms	25ms	50ms	75ms	Onsets
Scales	4	1	30	43	57	65	83	86	88
Schubert Lullaby	5	4	12	16	23	23	32	34	37
Seitz Concerto	14	13	21	34	44	52	66	70	78
Combined	23	18	63	93	124	140	181	190	203

Table 6.4: Number of false positive and false negative detected fingered pitch changes and correctly detected fingered pitch changes found within Xms of the hand labeled note onset. A negative time means the attack was detected prior to the label.

Results for note onset detection through bow change are given in Table 6.4. Real-time note onset detection using pitch change performs well with an overall F-measure of 91%. False positives and negatives were almost entirely due to lost contacts between the sensor conductive layers with occasional additional false negatives due to the player pressing the finger down gradually. Further refinements in sensor build may solve the minor remaining issues. 63% of notes are within 25ms and 56% meet our target of 10ms. Again, there are a high number of predictive onsets. The number of predictive onset labels is due in large part to off-the-string notes. In these cases, fingers are typically put down on the finger board prior to the bow. Notes with an off-string attack account for almost $\frac{2}{3}$ of onset detections more than 50ms prior to the expected onset. In these cases, attack onset should be preferred over pitch change because that corresponds better to sound production.

It was also possible to see a difference in onset detection timings between putting fingers down on the string versus lifting them up. The average difference in timing between hand labels and algorithm labels when adding fingers up the fingerboard is 30ms whereas removing fingers, the average difference was -25ms, regularly preceding hand labeled onset. This suggests that it takes 25-30ms to firmly contact the string. Note onsets with higher delays tended to fall into two cases: change of string allowing for late release of the finger on the previous string, or poor finger contact. Pitch changes with poor finger contact typically produced a period of time in which no clear pitch was evident in the spectral domain. In these cases, note onset detection tended to coincide with the point when the audio pitch became clean.

6.4.4 Note Onset through Slurred Repetition

Our tests only included two examples similar to slurred repetitions, these being re-attacks without changing bow direction, one of which is presented in Figure 6.6. In both examples the pressure reduction prior to re-attack caused the down force to go below the off-string attack algorithm's 0.44N limit for the bow remaining on string meaning both slurred repetitions were classified as an off-string attack even though the bow may not technically have left the string. With only two sample cases, one detected within 35ms of the hand labeled onset, and one labeled within 4 ms of hand labeled onset, it is not possible to draw robust conclusions, but onset detection techniques have so far demonstrated the ability to detect this often challenging case.

6.5 Conclusion

This case study has not only highlighted the potential usefulness of data from the augmented violin to help solve traditional music computation problems, but also demonstrates the effectiveness of our low latency low-cost bow tracking and pitch estimation techniques. We have demonstrated reasonable means for identifying note onset in violin performance through multiple performance cases: off-string attack, bow change, finger change, and slurred repetition. Though bow change can only be detected with reasonable success in post-processing, as shown in Table 6.5, off-string attack and finger change were effective in real-time. Within 25ms, a latency commonly used in research, detection of off-string attack has a precision of 96% and recall of 86% for an F-measure of 90% and detection of onset through finger change has a pre-

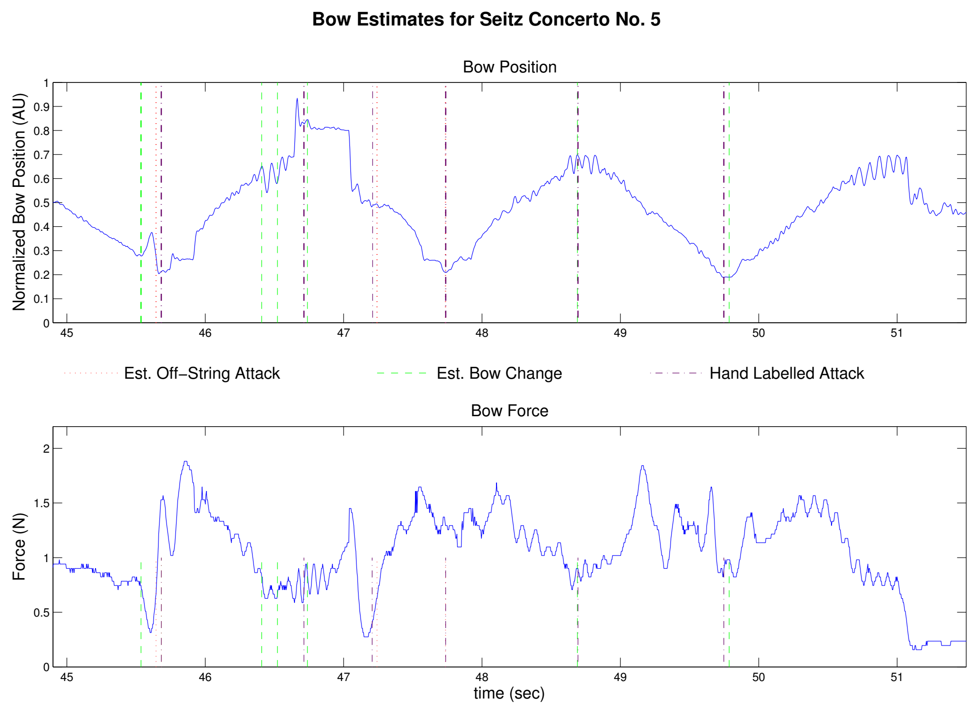


Figure 6.6: Bow-based note onset classifications for opening phrase of Seitz Concerto. Despite bow position estimate errors shortly before, slurred re-attack is easily detectable at 47.24s. An accented note is identifiable at 45.59s by the drop in force post attack along with the lift immediately required to re-attack the string. It is also clear that the tension metric derived prior to play for this session is too low as the clear drop in pressure after 51s suggests the bow is no longer on the string but the force estimate remains around 0.2N rather than 0N.

Onset Type		-50ms	-25ms	10ms	**25ms**	50ms	75ms	FP / P	FN / R	Labeled Onsets / F measure
Off-String Attack	#	20	27	76	**95**	107	109	4	1	111
	%	18%	24%	68%	**85%**	96%	98%	0.96	0.85	0.90
Bow Change	#	13	14	22	**30**	52	71	73	2	117
	%	11%	12%	18%	**25%**	44%	60%	0.29	0.25	0.27
Finger Change	#	63	93	124	**140**	181	190	23	18	203
	%	31%	45%	61%	**69%**	89%	93%	0.85	0.69	0.76
Slurred Rep.	#	0	0	1	**1**	2	2	0	0	2
	%	0%	0%	50%	**50%**	100%	100%	-	-	-

Table 6.5: Number of false positive (FP) and false negatives (FN) for different types of note onsets, including correct detections found within Xms of the hand labeled note onset. A negative time means the attack was detected prior to the label. Precision (P), recall (R), and F measure are calculated using onsets found within 25ms of the hand labeled onset. There are no F measure values (P, R, F measure) for slurred repetition as the sample set was too small for statistical value.

cision of 86%, and recall of 69% for an F-measure of 77%. Both these compare favorably with Böck's current off-line state-of-the-art. Both cases of slurred repetition we had were correctly detected, however the sample size is too small to derive any robust conclusions.

Chapter 7

Applications of the Augmented Violin

This chapter contains applications of the augmented violin and practical details of associated software. We present the real-time software VST that implements our methods of bow tracking and pitch estimation described in previous chapters. Subsequently we present applications that use the results from the real-time software, namely the custom pitch correction software used in the user studies in Chapter 8 and Chapter 9, and live audio visual performance applications.

7.1 The Low Latency Augmented Violin VST

While Chapter 4 on bow tracking and Chapter 5 on low latency pitch estimate provided hardware implementations and language agnostic algorithms for building a low latency augmented violin (LLAV), this section details a software implementation for real-time use. Real-time audio being one of the low latency augmented violin's core inputs, the software implementation was built as a Virtual Studio Technology (VST)[1] audio plugin using the JUCE[2] C++ libraries. The decision to build the LLAV software as a VST means it is portable across most major digital audio recording and editing software and enables us to utilize commercial audio processing capabilities. While MATLAB[3] was used for algorithm proof-of-concept, data post-processing, and bow training, versions of the LLAV VST were used in real-time data display, estimation and recording all data presented in this thesis. The LLAV VST consists of three major sections: a user interface (UI), a sensor handling process, and the audio run-time process.

7.1.1 User Interface

The main LLAV VST UI is shown in Figure 7.1. It contains various virtual hardware controls, and information about incoming sensor data and outgoing violin estimates. Virtual hardware controls are limited to selecting USB input and output ports (Figure 7.1 upper right) and turning hardware input and output in the software on and off. Software settings include choice of violin size, enabling/disabling sampling bow tension, switching between monophonic and polyphonic input operation, and selecting the size of the pitch analysis window (block size) and estimate frequency (hop size). The plugin generates data logs for post-processing which can be disabled or restarted through the UI (Figure 7.1 bottom right). Two tabs on the right side of the UI switch the interface to one for loading data logs enabling playback of a saved performance, and one for loading bow polynomials generated in MATLAB used for real-time bow position and pressure estimation (Section 4.5).

Data displays include a historical depiction of raw data from all four bow sensors and raw data from the fingerboard sensor for all four strings (Figure 7.1 left). Both these historical displays are useful for ensuring correct hardware operation and basic insight into sensor values. There is a similar historical display for the estimated pitch and also an instantaneous pitch display (Figure 7.1 bottom left blue box). The instantaneous pitch display displays the most recently

[1]The VST audio interface specification was created by and remains maintained by Steinberg: `http://www.steinberg.net/en/company/developers.html`. It is widely used for third party creation of audio effects and virtual instruments.

[2]`https://www.juce.com/`

[3]`http://www.mathworks.com`

Figure 7.1: The Low Latency Augmented Violin VST user interface. The interface enables setting hardware and software parameters and provides a graphic representation of both incoming data and outgoing estimates.

.

calculated hardware pitch estimate, $\hat{f}_{hw}$ from Equation 5.4, non-restricted pitch estimate, and the restricted pitch estimate. There are two more historical displays, not active in this example, for estimated bow position and bow pressure (Figure 7.1 upper right). Lastly, a general purpose output console exists to display text to assist debugging, confirm state change, or post additional data. In this example, it is streaming volume for the incoming audio. The UI updates every 100ms.

7.1.2 Main Audio Processing

The main core of the LLAV VST is the real-time audio processing thread. Besides VST housekeeping tasks, the audio processing thread is where pitch estimation occurs using the methods described in Chapter 5.1. The audio process is also where any data logging and all data output are coordinated.

The audio thread is called for execution by the audio host every audio processing block[4] Although we can change the block size to operate with lower latency, most of our real-time work uses a 512 sample block size (11.6ms with a 44.1kHz sample rate or 10.7ms with a 48kHz sample rate) as it yields better pitch estimates and is the smallest block that guarantees enough time for bow estimates to complete in the sensor processing thread. For pitch estimation the default is to use a four times overlap, i.e. a hop size of $\frac{1}{4}$ the block or 128 samples so that although we have a fixed 512 sample delay before result production, we get a pitch estimate result for every 128 samples.

As discussed in Section 5.7 we have two styles of pitch estimation: polyphonic audio input where each string has its own audio channel, and monophonic input where there is one audio channel for acoustic or electric violin audio. The VST will choose which form of pitch estimation to use based on the Mono/Poly setting in the UI. Additionally, the program supports both half-size and full-size violins, selectable through the UI. The fingerboard sizes are different and this changes the constants in Equation 5.4 to the appropriate dimensions.

[4]Though window size and block or buffer size are theoretically different, in this thesis, the VST block size and window size are always set equal and are functionally equivalent.

7.1.3 Data Output and Passing Parameters Between VSTs

We decided to make the LLAV VST a processing-only plugin so as to separate estimate generation from estimate effect. The LLAV VST is capable of sending output via USB-serial, OSC and as data encoded audio. Embedding the estimate results in an audio channel is the fastest, most consistent and predictable way to pass estimation results to another VST.

Along with passing the audio input channels unchanged, the LLAV VST creates additional data as audio channels. In polyphonic mode, a data as audio channel is created for each string's pitch estimate result, along with a combined monophonic output. In monophonic operation only one data encoded audio channel is created. The data encoded as audio, given in Table 7.1, was chosen as it represents the range of the data produced which may be used elsewhere. The choice to always send the volume for each string, but not the pitch estimate was based on the assumption that acting on a single audio channel, the relative volumes will imply whether a the pitch result is interesting in a polyphonic context, but the pitch estimate from another string is not interesting without the corresponding audio. The polyphonic combined data channel uses the pitch estimate corresponding to the one from the string with the highest RMS, and is expected to convey the primary monophonic melody.

A sample routing example for the LLAV is shown in Figure 7.2 using the retuning plugin presented in the next section as estimate recipient. Each data as audio channel sends:

The 16 floats in Table 7.1 are followed by another 16 floats-worth of zeros. These 32 floats are repeated for a full hop. The next hop is composed of the 16 floats derived from the next pitch estimate with 16 zeros repeated and so on. The data is placed in the audio buffer and all of the up to 5 channels of audio data can be routed just like any audio channel. Along with marking the beginning, end, and internal order of the packet, the fixed values enable a check at the receiving end in the event the gain on the data audio channel has been changed in the host VST. It is common and easy in audio software to alter a channel's volume[5] and the fixed values throughout the packet, enable both identification of a valid packet, and the amount to rescale to return the values to the original.

Data encoded in the audio channel is retrieved and validated in the receiving VST. The data

[5]In our audio host, Reaper, the data as audio had to be rescaled in order to not trigger auto-muting

Data Type	Description	Value	Num. Floats (8 bytes)
Packet Begin	Fixed	1.0	1
Time Stamp	Time since start, rolls over above a fixed max.	0 - 2^{29}	1
Channel Num. Begin	Fixed	-0.9	1
Channel Num.	Which channel is pitch est. for? (4 is combined)	0 - 4	1
Pitch Estimate Begin	Fixed	-0.8	1
Pitch Estimate	Sensor informed pitch est. for given channel	0 - 4400	1
String Vol. Begin	Fixed	-0.7	1
String Vol. (RMS)	Volumes (RMS) for strings G-E	0 - 2	4
Bow Position Begin	Fixed	-0.6	1
Bow Position	Between 0 (frog) and 1 (tip)	0 - 1	1
Bow Pressure Begin	Fixed	-0.5	1
Bow Pressure	Grams of force measured	0 - 1000	1
Packet End	Fixed	-1.0	1

Table 7.1: Format of data packets encoded on an audio channel. Data types marked as *XXX Begin* are fixed values to help check validity of the packet at receipt.

audio channel is broken up to look for a valid packet of data in each hop. If two opposite samples are found 16 samples apart, the entire provisional packet is scaled so the start and end are 1.0 and -1.0 countering any volume effects. Internal packet fixed values are checked and if they match Table 7.1, the packet is deemed valid and the encoded data can be pulled out for use by the recipient VST.

Example Use of LLAV VST with Routing

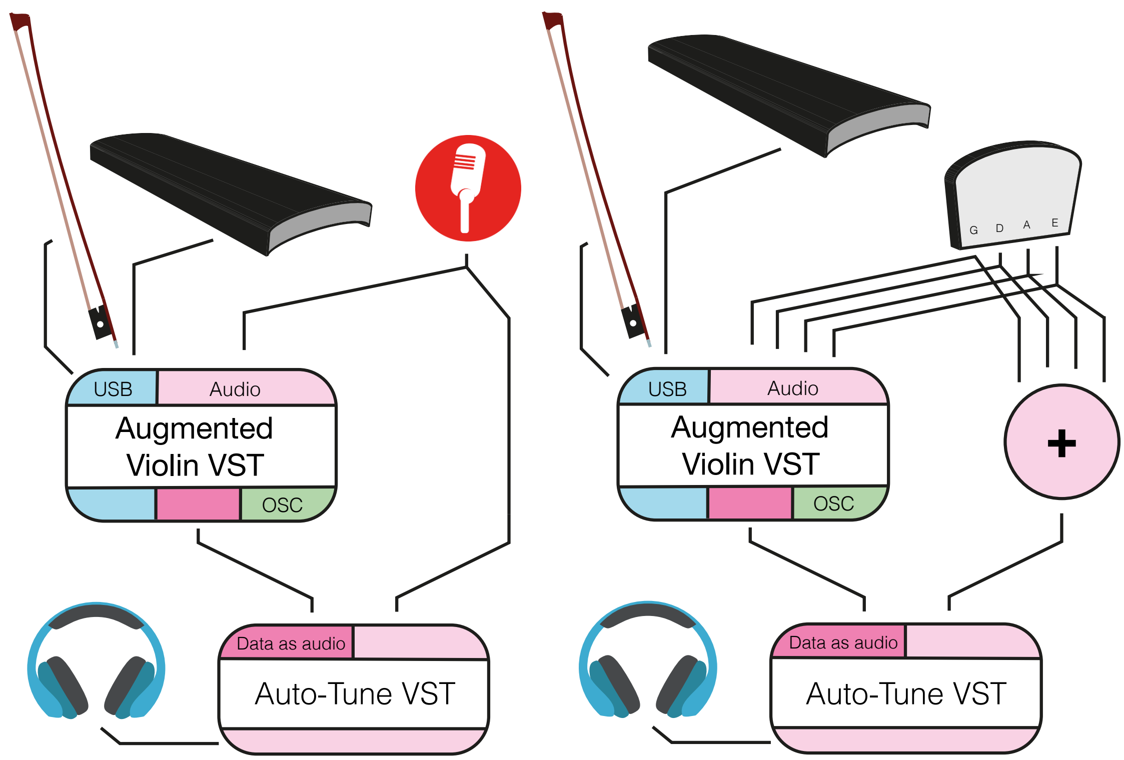


Figure 7.2: Examples of routing audio and data for single audio source (left) and polyphonic bridge source (right) for the augmented violin VST used with an retuning VST presented in Section 7.2. Input and output modes are color coded: USB is blue, OSC is green, audio is pink with data encoded as audio is darker pink.

Figure 7.2 shows examples of using the LLAV VST with the retuning described in Section 7.2 in both monophonic and polyphonic modes. The figure includes basic information on routing audio and data between hardware input sources, namely the augmented bow and fingerboard, and the audio source. In both, sensor data input from the fingerboard and/or bow is received over USB within the LLAV VST where the hardware pitch and bow tracking estimates are derived.

Starting with the example using a single audio source (Figure 7.2 left), input generated from a microphone or pick-up is routed via an audio channel in the audio host to the LLAV VST where the hardware estimates and audio signal are used to produce the restricted pitch estimate. Analysis results are sent as data over audio to an retuning VST (Section 7.2) where the original audio input is retuned using the LLAV VST pitch estimates. In the polyphonic example (Figure 7.2 right), the overall data flow is similar, though in the polyphonic case, each string's audio channel is routed separately to the LLAV VST. Additionally, this example uses the retuning VST from Section 7.2, which expects a single audio input so polyphonic audio must be mixed.

7.1.4 Sensor Data Processing & Real-Time Bow Tracking

Sensor data processing is performed on its own thread allowing faster data sampling than would be possible through the audio process thread if using window sizes over 128 samples. Sensor data processing consists of retrieving USB-serial data from the remote AVR hardware controller in Section 4.3.3 and Section 5.3, basic parsing of the received data and computation of real-time bow tracking estimates. Data processing is enabled/disabled through the UI's virtual hardware on/off button.

Real-time estimation of bow pressure and position requires both the polynomials describing the bow's behavior and the bow tension for the performance session. The system must be trained for the bow being used with characteristic polynomials derived using MATLAB as described in Section 4.5. Characteristic polynomials are loaded through the *Bow* tab selected through the main UI and tension is sampled using the UI in Figure 7.1.

As per Section 4.5.2, we estimate a subset of expected sensor readings for any given combination of bow weights and positions and then compare the momentary sensor readings to the set of expected readings. The process is computationally intensive, especially considering characteristic polynomials are typically higher order. In order to reduce the per estimate computation, we calculate the set of expected sensor readings once, immediately after any change in sampled tension or characteristic polynomial, and use the resulting matrix for look-up when making a real-time estimate. Further, code is vectorized to improve runtimes.

We are currently running a version of bow tracking which estimates the expected sensor reading for each sensor every 3.25mm along the full length of the bow and every 2g of down force up to 400g. This resolution makes for a look-up matrix for each sensor with 40,000 elements. With four sensors, this results in 160,000 points evaluated each estimate. Compiled without compiler optimizations, it takes on average 4.55ms to calculate a bow estimate, or 0.18 ms when compiled to optimize for speed.

7.1.5 Augmented Violin Playback

Along with real-time functionality, the LLAV VST includes a playback function enabling a previous performance to be reproduced using the same data inputs as when performed. Playback requires the sensor log recorded by the VST when originally performed and the originally recorded audio. Data from the log is synchronized using the audio playback sample count from the audio. Sample count is included in the data recorded to the sensor log during recording.

Playback allows altering parameters elsewhere in the audio chain with the same repeated test inputs, making it helpful in debugging. It also enables replaying user performances to validate and evaluate performance events.

7.2 Retuning VST

In order to accomplish our investigations into aides for learning intonation and experimenting with the effects of pitch simplification, we built a retuning VST. Both our ideas for providing an aural guide with correct pitch, and making pitch easier rely on pitch correcting performed audio. Using the low latency, high-accuracy pitch estimates from the LLAV VST enables our retuning VST to perform low latency automatic correction of the augmented violin.

Auto-tuning, commonly used when producing vocals in pop music [167, 84], is a technique where a recorded performance is analysed for performed pitch and then automatically shifted to be in tune while still retaining the original audio quality. Pitch is altered without altering timing. Rather than start from scratch, we modified Tom Baran's open-source *Auto-Talent* plugin[6] to accomplish our version of auto-tuning. We ported Baran's algorithm to the JUCE/C++ environment for subsequent compilation to a VST plugin and modified the code to use the low latency pitch estimate from the LLAV VST. As discussed in Section 2.5, we also wanted to experiment with different styles of pitch correction for more nuanced styles of pitch snapping.[7]

The basic structure of the Auto-Talent algorithm is

1. Estimate performed input pitch.
2. Calculate target pitch.
3. Shift audio from input pitch to target pitch.

We left the actual pitch shifting algorithm in place, but replaced the first two stages, with the pitch estimate in the first step coming from an external VST, currently the LLAV VST, and rewriting target pitch calculation to meet our goals for controlling the degree of snap. Auto-Talent uses a time-domain overlap-add technique [170] that is synchronous with the pitch period of the input pitch to shift the output pitch to the target pitch. All other aspects of the original Auto-Talent plugin were ported or replicated with the addition of a simple user interface, shown in Figure 7.3, for accessing all parameters.

Original Auto-Talent

The original Auto-Talent plugin estimates pitch by using auto-correlation. The pitch is then linearized prior to calculating the target pitch. Pitch is converted from the estimate made in

[6] `http://tombaran.info/autotalent.html`

[7] *Pitch snap* refers to the process of pulling a pitch to the nearest selected scalar pitch.

Hz to the linearized scalar version based on cents. The linearized scalar version of the input, $\hat{f}_{LI}(n)$, uses:

$$\hat{f}_{LI}(n) = -12log_2(\frac{440\text{Hz}}{\hat{f}_{in}(n)}) \tag{7.1}$$

The function is centered so that 440Hz[8] is zero and each semitone change has a value of one. Giving each semitone change a value of one means that every integer is a different chromatic note. The user is able to change the center frequency by changing the 'Concert A Ref.' in the UI.

Target pitch is determined based on the scale selected by the user. If the user has made no scale selection, the target pitch is the closest note in the chromatic scale, otherwise it will be an octave variant of the equal tempered notes in the user selected scale. For an example, in Figure 7.3, the estimated linearized pitch is 1.76 (487Hz) which rounds to B4, two semitones above A4. As 1.76 is below 2.0, we say the pitch is 24 cents flat of B4 and the fully snapped target pitch is 2.0, B4. Fully snapped outputs are theoretically always in tune.

The Auto-Talent plugin also allows two alterations of the target pitch. The first is shifting audio pitch by adding a pitch shift value to all target pitches through the 'Pitch Shift' UI control. The second is through the UI's 'Correction Weight' which will multiply the amount to shift by the selected weight. Once target pitch is calculated, pitch shift is carried out using the time-domain methods described earlier.

Other functions remaining from the original Auto-Talent are formant filtering which removes formant content during pitch shifting, and 'mix' which lets the user decide the gain balance in the audio output between unshifted audio and shifted audio.

User Interface

Figure 7.3 displays the user interface for the retuning plugin. It allows control of the expected central pitch, notes in the scale, whether to exclude any detected formants during pitch shifting, the ratio of how far to pull the pitch towards the target pitch (Correction Weight), application of a universally applied pitch shift (Pitch Shift), and the ratio of original audio to

[8] We are using equal temperament and a 440Hz A as our definition of correctly pitched.

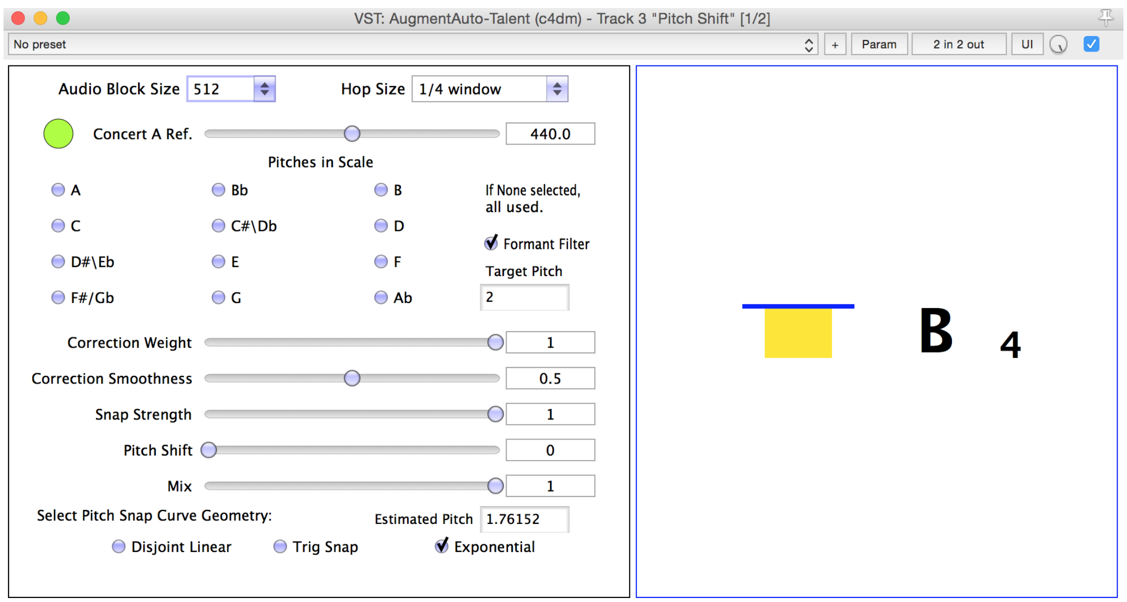


Figure 7.3: Augmented Retuning UI: Allows the selection of pitch snap properties and notes included in the snapped scale. The user can also see a both the estimated pitch and target pitch. The version depicted here also includes a graphics window showing the performed pitch and the relative intonation.

shifted audio (Mix). These parameters were all available in Baran's original Auto-Talent. We have added speed of snap application (Correction Smoothness), degree of snap applied (Snap Strength), and options for snap shape. We also enabled the user to select both the processing block size and the hop size separately of the audio host.

7.2.1 Using the Sensor Assisted Low Latency Pitch Estimate

The Retuning VST plugin takes two audio channels in: one of regular audio, one of data encoded as audio according to Table 7.1. Data from the audio channel is extracted as described in Section 7.1.2 and the pitch estimate from the LLAV VST is used as the current momentary pitch. Although our retuning plugin is capable of using a smaller or larger block-size, changing latency, we typically used a 512 sample block for a quarter of the latency of the the original Auto-Talent plugin, which uses a 2048 sample block.

With the estimate of the performed pitch, the next step is to find the pitch target which is the pitch we would like the audio shifted to. Pitch is first linearized as with the original Auto-talent. For determining target pitch, we added additional snapping variations according to Section 2.5. This included adding two new shapes of pitch curve, trigonometric, and exponential, and a strength setting to control the shape of the three pitch shapes.

Calculating Pitch Curves

The equation for the pitch curve for linear segments, where ψ is the strength factor, between zero and one, and p_i is the fractional pitch input between zero and one with zero being a note on the chromatic scale and one being the semitone above:

$$p_o = \begin{cases} (1-\psi)p_i & \text{for } p_i <= 0.5 \\ (1-\psi)p_i + \psi & \text{for } p_i > 0.5 \end{cases} \tag{7.2}$$

and the equation for the trigonometric curve is based on an arctan:

$$p_o = \frac{\arctan(20\pi\psi(p_i - 0.5))}{2\arctan(10\pi\psi)} + 0.5 \tag{7.3}$$

Lastly, the exponential pitch curve is described based on the following equation:

$$p_o = \begin{cases} -0.5 \times (|2 \times (p_i - 0.5)|)^{(1-\psi)^2} + 0.5 \text{ for } p_i <= 0.5 \\ +0.5 \times (|2 \times (p_i - 0.5)|)^{(1-\psi)^2} + 0.5 \text{ for } p_i > 0.5 \end{cases} \tag{7.4}$$

Referring back to Section 2.5, Figure 2.10 uses these three equations with a strength, ψ, for all three curves of 0.5. Figure 2.11 was drawn using the exponential pitch curve above as well.

Snap smoothness was also added to the VST which delays the onset of the pitch snap. For speed of snap application, with p_o as pitch output, p_t as the snapped pitch target, and p_i as the as played input pitch, within both the VST and Figure 2.12 we use:

$$p_o = (1 - (1 - \nu)^{t_s})p_t + (1 - \nu)^{t_s} p_i \tag{7.5}$$

Regardless of whether the audio is shifted, all audio is passed through the retuning VST. As a result, if the pitch estimate is momentarily significantly off, it may cause audible glitches even when not snapping the output pitch. Minor or stable errors in the pitch estimate will not cause the non-snapped audio to glitch, but will affect the snapped versions. Additionally, if a pitch estimate is stable but inaccurate, the audio will be shifted an incorrect amount making the snapped output out of tune.

7.2.2 Visual Pitch Feedback

As can be seen in Figure 7.3, the retuning UI has a second window which shows live feedback on the pitch being performed. The UI shows the standard note name of the note played, A0 being the lowest note on a piano and A4 being 440Hz, and a visualization depicting how in tune the performed note is in relation to the nearest chromatic note. A rectangle grows from the centered bar with size and color determined by the difference between the performed pitch and the displayed equal temperament note. If the pitch played is sharp, the rectangle grows above the center bar, and if flat, below. The color is also set on a gradient between green and red where green is low or no difference through yellow and orange to red if the note is highly out of tune. In order to reduce flicker within the visualization, the pitch visualized uses an FIR filter weighting more recent pitch estimates more heavily:

$$\hat{f}_{vis}(n) = \frac{2}{N(N+1)} \sum_{i=0}^{N-1} (N-i)\hat{f}(n-1) \tag{7.6}$$

We used $N = 20$. If there is no audio the intonation visualization is blank with only the center bar remaining.

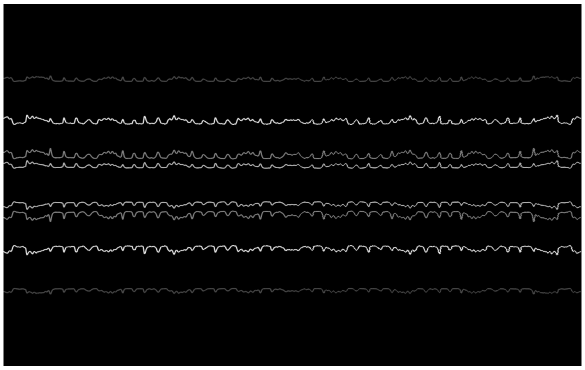
7.3 Live Performance

Though functionality in the context of learning and complexity management is the primary target for augmented violin research, the instrument capabilities can be used in other applications. We used the augmented violin in live performance making use of the real-time sensed outputs to control visuals in an effort to enhance audience engagement in classical performance. We currently have two very different visualisations.

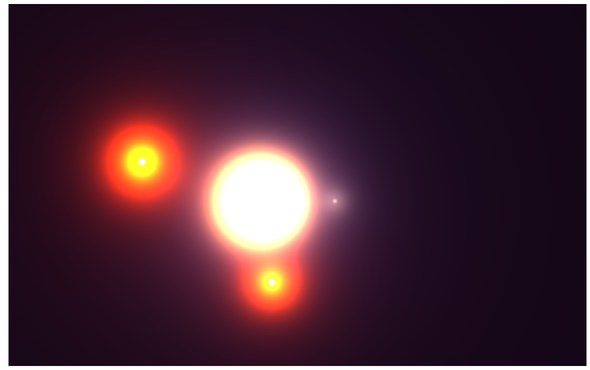
7.3.1 Computer Visual Music Response

The author teamed up with interactive music performance systems expert Adam Stark[9] to create computer graphics driven by the augmented violin outputs. Two visuals, depicted in Figure 7.4, were created using Cinder[10], a C++ library for visualizations. Visualizations are controlled by sending output from the LLAV VST to the visualization software through OSC. The use of OSC means the visualizing computer can be remote, often useful in performance environments. It also means that the computational load can be split. The two visualisations use raw bow data along with the pitch and volume estimates from the LLAV VST.

[9] `http://www.adamstark.co.uk`
[10] `https://libcinder.org/`



(a) Bow History



(b) Audio Orbs

Figure 7.4: Computer controlled visuals driven by real-time sensor data from the augmented violin. a) depicts raw bow sensor history mirrored around the middle of the screen. b) consists of floating orbs with size controlled by violin volume and color controlled by the momentary performance pitch class.

These visuals were premiered at the closing party for Digital Shoreditch 2015 and have been seen in repeat performances since. The author played selections from Bach's Sonatas and Partitas on a 1903 J.P. Ditter violin equipped with the augmented fingerboard and an augmented bow.

7.3.2 Stage Lighting Control

A second real-time augmented violin controlled visual performance was to control stage lighting based on bow pressure and string played. Using the polyphonic violin, the collaboration with Toby Harris involved four par can lights each driven by the RMS of an assigned string. A single additional spotlight had brightness controlled by estimated bow pressure. The lights were controlled through DMX by an embedded lighting controller designed and programmed by Harris. The light controller received augmented violin performance estimates from the LLAV VST through virtual serial over USB.

These visuals were also premiered to a large audience at the closing party for Digital Shoreditch 2015. The author performed Arvo Pärt's *Fratres* with piano accompaniment using the

Figure 7.5: The author controlling stage lights through the augmented violin.

polyphonic electric augmented violin with the augmented bow.

7.3.3 MIDI Violin

For the 2015 Mathematics and Computation in Music Conference, the author was requested to use the augmented violin to perform Gareth Loy's *Blood From a Stone*, originally written for Max Mathews's electric violin. The piece uses a mixture of the acoustic violin performance and electronics sounds triggered by data from the violin. Loy has rewritten the code for *Blood From a Stone* to use MIDI input from a MIDI violin. As our augmented violin already offered sensor-rich performance data with low latency, it was possible to covert it to a compatible MIDI violin.

The LLAV VST was adapted to issue MIDI note ons and offs based on performance data. As development of the MIDI violin so far has been for a one-off concert, for programming expedience, MIDI note commands were determined by watching for note onset from changes in the pitch estimate or RMS volume rising above (or below) a threshold. The overall performance of the MIDI violin was satisfactory though it was prone to reporting multiple note ons (onsets) for a specific note if the transition between notes was imperfect. As spurious note ons were typically immediately followed by a note off, it was not problematic except in cases where the software for *Blood From a Stone* spawned multiple new events for every note. Still, 95% of the piece was successfully performed with the augmented violin acting as MIDI violin.

7.4 Conclusion

In this chapter we presented the practical implementation of the LLAV VST augmented violin software, the introduction of a pitch retuning VST, and three different performance interactions using the augmented violin. Section 7.1 introduced the LLAV VST which implements the algorithms from Chapter 4 and Chapter 5 to produce the bow tracking and pitch estimates based on sensor and audio input. These estimates can be forwarded through OSC, MIDI, USB-serial, or a novel technique for encoding data on audio channels for use in further applications.

We presented four applications driven by data produced by the augmented violin. The first is a retuning VST allowing us to correct pitch in real time and designed to test our ideas from Section 2.5 on complexity management in relation to pitch. The retuning VST lets us perform pitch correction with selectable shape, strength, and speed which we use and test extensively in Chapter 8. The retuning VST also features heavily in Chapter 9 providing the aural guide through pitch corrected audio and visual feedback on pitch.

The three other applications discussed were all real-time live performance applications that have been used in public performance. Two used performance data from the augmented violin and LLAV VST to drive live visuals highlighting the performer's musical actions through projected computer graphics or programmable par can stage lighting. The third was an implementation of the augmented violin as a MIDI violin for performing Gareth Loy's *Blood From a Stone* which uses performed notes to generate new sounds beyond the violin's normal capabilities. These three performance cases demonstrate the capability of the augmented violin to function not just as a study tool, but also as an interesting and rich source of new performance opportunities.

Chapter 8

Expert Pilot Study

The design and build of the augmented violin was motivated by the desire to explore the use of technology to assist violin learning, and in particular, look at real time feedback to assist intonation learning and whether we can use technology and complexity management to simplify the violin theoretically improving a beginner's experience learning violin. To investigate these ideas we conducted two user studies. The first study, the expert pilot study presented in this chapter, was intended to test the augmented violin's usability along with laying initial ground work for investigation of our second objective, simplifying beginning violin performance. We carried out a pilot study with advanced players focusing on the physical feel and audio quality of our augmented violin and retuning VST. At the same time, we tested out ideas on pitch simplification studying the effects and experience of automatic pitch correction. Results were used to inform our second study with beginner students, presented in Chapter 9, where we investigated the use of the augmented violin as a learning aid within a lesson context.

8.1 Study Motivations

The first goal of the pilot study, verifying physical playability of violin augmentations, was key to validate whether it meets core design objectives that the augmentations be sufficiently accurate, and as seamless as possible so as not to interfere with play. Also, as pitch simplification relies on altering sound, we needed to ensure acceptable audio quality. If the augmentations are too inaccurate or impede the ability to play the violin normally, we would not expect any user, beginner or experienced, to want to play the augmented instrument for any extended period of time. Additionally, the closer the feel of the violin and bow augmentations to a regular violin, the better the skills transfer; skills learned on an unaugmented violin can transfer with minimal adjustment on the augmented violin, and vice versa. It is also important to ensure that practice with the augmented violin will not inherently encourage bad technique. As Johnson [76] says, if a learning aide requires or conceals poor technique, it might be significant work for a student to correct that poor technique when switching back to normal play. Usability tests are a must.

The second goal was to evaluate different forms of automated pitch correction, partially, or fully eliminating pitch error. One of the core principles necessary for complexity management to be successful in the context of the augmented violin is that all learning on the augmented violin and associated systems must contribute to technique applicable on a non-augmented traditional violin. As discussed in Section 2.5, making performance on the augmented violin theoretically easier by snapping notes so they are always in tune not only eliminates expressivity, but we also expect it to interfere with the normal aural feedback a violinist uses to evaluate and correct pitch.

The prevailing view in neuroscience literature, as presented in the widely cited work of Zatorre, is that,

> Feedback interactions are particularly relevant in playing an instrument such as a violin, or in singing, where pitch is variable and must be continuously controlled. The performer must listen to each note produced and implement appropriately timed motor adjustments. If auditory feedback is blocked, musicians can still execute well-rehearsed pieces, but expressive aspects of performance are affected [190].

This has created the idea that "auditory feedback plays a minor role in error monitoring," [104]. However, looking deeper, the research referenced is restricted to the piano [146, 43, 150, 138], a percussion instrument where basic tone and intonation are innate to the instrument.

In contrast, when Chen studied shifting in cellists, [25], aural feedback was demonstrated to be critical for shift intonation. Beyond adding depth to research about the effects of removing aural feedback in instruments other than the piano, we also want to confirm the effects on intonation when doing even the most basic violin performance, not just shifting. Playing without shifting requires only motor memory of finger placement, not motion of the whole left arm meaning we are not guaranteed to obtain the same results as in Chen's study.

While fully corrected pitch may simplify the requirements for early success faced by a beginner, it is likely full pitch correction would impede learning overall. Pedagogy suggests, as discussed in Section 2.4, good intonation is a learned skill requiring accurate error identification and correction. Assuming fully corrected pitch does impede the ability to play in tune and correct if wrong, we ask if there is a way that we can improve the intonation on pitch output heard by a beginner without eliminating their ability to improve. We created the variable pitch retuning VST in Section 7.2 in order to give us the ability to vary the degree and style in which we correct pitch so we can explore the effects of different levels of pitch correction as a means of pitch simplification. Within the pilot study we wanted to see the potential effects of the different pitch snaps on performance in order to learn:

List 2 Research questions on different forms of automated pitch correction.

1. whether limiting aural feedback by removing pitch error around a particular note (full pitch correction) will negatively impact performed pitch.
2. whether altering aural feedback by reducing, but not eliminating pitch error around a particular note (partial pitch correction) will negatively impact performed pitch less than full error removal.
3. whether gradual pitch correction (delaying reduction of aural feedback) will negatively impact performed pitch less than instantaneous full error removal.
4. whether study participants have a aural preference between different shapes of pitch snap curve and speeds of pitch snap.

The fourth question in List 2 differs from the first three in that it is predominately an experiential question. Though the third question focuses on whether a slow speed of pitch snap may enable a violinist to hear enough of the performed pitch to begin correction, as mentioned in Section 2.5, shape and speed of snap may directly effect sound quality. Discrete changes in snap direction or rapid snap may feel and sound artificial. We would like to know if there is a perceived difference and if so, what sounds best.

Lastly, for both our goals, investigating usability and pitch correction, we want study participants who are likely to notice and react to differences in either physical violin performance or pitch snap. Beginners are less likely to be intimately familiar with how to use the bow correctly, or play in a way that makes demands on the instrument. Similarly, as beginners often have a weak understanding of intonation and have not yet learned to quickly correct pitch, in a short study, experts can be expected to respond more strongly and knowledgeably to alterations in the intonation process. Experts will be more sensitive to audio delays, alterations to play due to sensors, and unreliable sound quality. Expert players will be better able to explore what the system is doing, and are more likely to provide consistent performance which will let us verify the system is performing as expected. For these reasons, we chose to conduct the pilot study with expert violinists.

8.1.1 Expectations

As the author is an expert violinist, we felt reasonably confident that the fingerboard augmentations to the violin would not interfere with regular play or require significant change in technique. The augmented bow is roughly 10g heavier than the traditional bow and the cabling, chosen to maintain largely normal bow balance, is indeed annoying but tolerable in our experience. Still on the whole, the bow is expected to be acceptable and not interfere with most bow techniques even if it may take some time to get used to.

We actively wanted to highlight and discover flaws with the augmented violin sound and feel and hence, the requested feedback focuses on asking participants about qualities they did not like. Additionally, we expected classically trained experts to be tied to both the specific sound and feel of the acoustic violin when we were using an electric. Still, we are not asking them to like the system but to verify it is of sufficient quality and reliability to use with beginners.

For pitch snapping, based on the importance of aural feedback to the intonation process loop and evidence from Chen's study [25] that cellists used aural feedback for correction, we expect violin to differ strongly from piano performance, and hypothesize that even with minimal shifting, advanced players will still have poor intonation without aural feedback. If true, we hypothesize that restoring some error should improve results through both partial snap and speed of snap though partial snap is the more likely of the two as any pitch snap fast enough not to be distracting may not enable enough time for aural feedback to impact performance.

In terms of personal enjoyment and preference, expert violinists should have a developed pitch process loop which is somewhat subconscious. While they will be better able to notice and respond to intentional alterations in pitch snapping, they are more likely than a beginner to find loss of subtle pitch control confusing and distracting. As Johnson [76, p.44] says,

> While working on one aspect of playing [learners] will also be relying on automated processes they have learnt previously. As the research on attention and skill shows, drawing attention to an automated process can damage performance. Therefore it is important that real-time feedback is only used to mediate attention towards aspects of technique that are in the stage where they are being consciously learnt.

Further, a professional is used to using vibrato, portamento, and may intentionally choose not to follow equal temperament tuning. Because of this, we expect unease at the loss of expressivity and disruption to their internalized link between physical input and audio output. While we expect that snapped versions of audio will be objectively more in-tune, we do not necessarily expect that players will enjoy it because the snap does not align with their extensive experience.

We hypothosize that participants will prefer a pitch curve shape where the transition between notes is smooth and continuous and is not disjoint as with the linear segmented snap. We expect performers to dislike the instantaneous snap for sounding artificial but dislike an overly slow snap, as the the pitch might noticeably change without any obvious input on their part.

8.2 Expert Pilot Study Design And Execution

The expert pilot study asked 8 experienced violinists to play the augmented violin with retuning software described in Section 7.2 and provide feedback during a single session. The session was split into an familiarization section and three test sections, each testing a different pitch correction factor: *shape*, *strength*, and *speed*, as presented in Section 2.5 and detailed in Section 7.2.1. Each test section was composed of nine trials, with two trials in the familiarization section. In each trial, participants were asked to play a short segment from a song and answer the same four questions (List 3) relating to their experience of pitch using the Likert scale. At the end of each section, there were additional questions about the overall audio and pitch correction within the section, and towards the end of the study once all trials were complete, there were questions about the functionality of the fingerboard augmentation, the bow augmentations, and the overall system performance and experience. It was expected each user should be able to complete the full 29 trials and survey questions in 60-90 minutes.

This study used the electric augmented violin with a polyphonic bridge described in Section 5.7.1 along with the augmented bow described in Chapter 4. The Low Latency Augmented Violin VST (LLAV VST), with sensor input, was used to estimate pitch which was then used by the retuning VST to pitch snap the audio heard by participants. Participants were asked to wear a pair of passive sound isolating headphones, Sennheiser HD25 Mark 2s, so that they would not be able to hear any acoustic sound from the electric violin, but only hear the test condition, a variably pitch corrected version of themselves.

We measured overall latency of the system at 37.6 ms. Latency was measured by utilizing a Bela board to generate an audio sine wave routed to the host computer, a 2014 Macbook Pro, using a Behringer FCA1616 USB audio interface (I/O). The sine wave was then processed in Reaper, set for a 256 sample audio input buffer, passed through the LLAV and pitch shifting VSTs before being sent back out the audio interface where we measured the delay between starting to send the sine wave and first receiving it at the (I/O) headphone out jack. Running the sensor input did not have any impact on overall latency. We confirmed that changing the size of the VST processing block impacted the latency in-line with the block length,[1] i.e.

[1]To clarify, there are two buffer sizes. The first is the audio input buffer size for Reaper retrieving and sending audio to/from the audio interface. The second is the VST processing block size which is the size of

Section	Trials	Option	Shape	Strength ψ	Speed ν
0) Familiarization	1-2	**1**	**Exp.**	**0.50**	**0.0**
		2	Linear	0.90	1.0
1) Shape	3-11	**1**	**Linear**	**0.00**	**0.8**
		2	Linear	1.00	0.8
		3	Exp	0.99	0.8
2) Strength	12-20	**1**	**Exp.**	**0.00**	**0.8**
		2	Exp	0.50	0.8
		3	Exp	1.00	0.8
3) Speed	21-29	1	Exp.	1.00	1.0
		2	Exp	1.00	0.5
		3	**Exp**	**1.00**	**0.0**

Table 8.1: Study structure and settings. Each section had three options, each trialled three times in random order within the section. Options along with their settings in bold combine to act as no pitch shift or the instrument heard as played.

with our default VST block size of 512, the inherent system delay receiving and sending audio without any VST processing is 26.0ms. The system delay is obviously much larger than the 10ms target latency and can be reduced slightly by reducing the Reaper input buffer size to 128 samples (31.2ms of which 19.6ms, is due to audio conversion and buffering), however the lower latency was accompanied by unacceptable audio artefacts.

8.2.1 Sections of Study

The three test sections, Section 1-3 in Table 8.2.1, were considered independent of each other, each testing three variations of a given pitch correcting parameter. Within each section each test case was used three times. For example, during the section testing strength, there were three values of ψ tested and each value was used in three trials, for a total of nine trials. The order in which the variations were encountered was randomly determined[2] and neither the

the data block being analyzed.

[2]For repeatability, each user was assigned a predetermined random number generator seed.

ID	Composer	Piece	Bars	Edition, Arr
1	JS Bach	Minuet in G Major (BWV Ang. 114)	17-32	Fischer, Seely-Brown
2	JS Bach	Gavotte in G Minor (BWV 822)	1-17	Fischer, Seely-Brown
3	JS Bach	Minuet in G Major (BWV Ang. 116)	1 - 32	Fischer, Seely-Brown
4	Brahms	Lullaby (or Cradle Song Op. 49, No 4)	All	Moffat
5	Seitz	Student Concerto No. 5, 3rd Mvmt	12-47	Summy-Birchard Inc.
6	Becker	Gavotte	1-17	Summy-Birchard Inc.
7	Ponce	Estrellita	1-8	Belwin Mills, Halle
8	Brahms	Hungarian Dance No. 5	1-32	Belwin Mills, Halle
9	Tschaikovsky	Marche Slave	1-16	Belwin Mills, Halle
10	Beethoven	Minuet in G Major	1-24	Belwin Mills, Halle
11	Schubert	Serenade	1-24	Belwin Mills, Halle
12	R. Schumann	Traumerei	1-16	Belwin Mills, Halle

Table 8.2: Musical excerpts in the study. Excerpts 1-6 are taken from Suzuki violin repertoire. Excerpts 7-12 were taken from *52 Masterpieces for Violin & Piano in First Position*. Participants were instructed to ignore all double-stops and were free to use any fingering and bowing they chose.

participant nor the author were shown which curve variation was being tested. Additionally, each trial had a randomly assigned musical excerpt the participant was required to play. Musical excerpts were taken from a pool of 12 song segments given in List 8.2.1, all of which can be found in Appendix A.

Musical excerpts were chosen for being easily sight readable, used at least a full octave, and were expected to take between 25-45 seconds to play. Only Excerpt 4, Brahms Lullaby, required any shifting. Within the song selection, there was an attempt to cover the basic range of a violin, from the G string to third position on the E string (MIDI notes G3 - D6) and cover a range of bow strokes, from *legato* to *staccato*. The decision to select 12 musical excerpts was intended to prevent participants becoming overly familiar with an excerpt by preventing them from playing any segment more than three times, while also having songs repeated enough to allow comparisons within and between players.

In the first section testing different shapes, users were given the pitch curves in Figure 8.1: (a) a linear pitch curve where pitch out matched the pitch in, (b) linear segments where an input pitch is snapped to the nearest in-tune chromatic semitone, and an exponential curve at nearly full strength. The target was to see how pitch snap shape impacts player performance and whether there were any preferences, particularly between the linearly segmented snap vs. the exponential snap (question 4, List 2). The no snap case was required as a control and the other two snaps were chosen to be strongly snapped in order to deviate farther from the control.

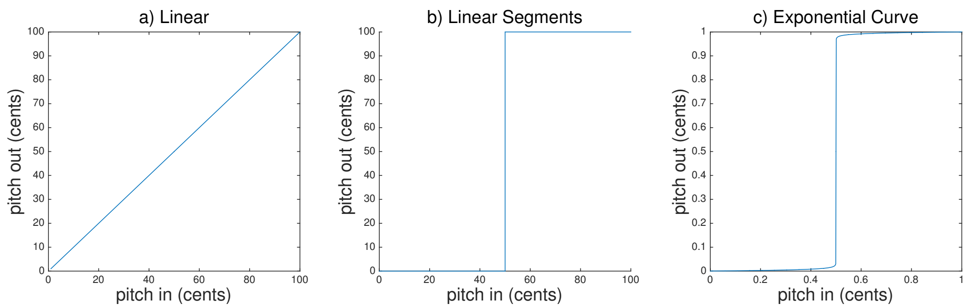


Figure 8.1: Three tested shapes of pitch input to output curves. (a) is the violin's normal interaction where input matches output, or a linear segment with strength $\psi = 0$. Shape (b) is a full snap performed using the linear segment with $\psi = 1$ and (c) is a strong exponential $\psi = 0.99$.

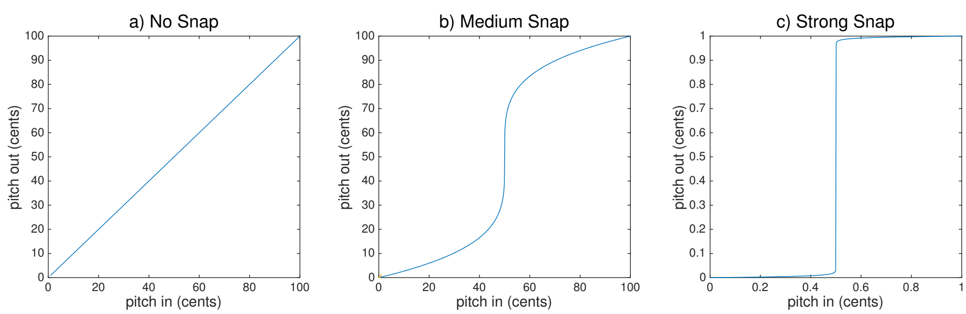


Figure 8.2: Three tested strengths of pitch input to output curves using an exponential shape. (a) is a low snap exponential with strength $\psi = 0$ where input matches output. (b) is a medium exponential with $\psi = 0.5$ and (c) is a strong exponential with $\psi = 0.99$.

The second section was designed to test the impact of pitch snap strength. We wanted to see how performance and preference changes between no snapping, strong snapping, and a snap in the middle that pulls towards the correct pitch but still allows heard error. As featured in questions 1 and 2 from List 2, we are particularly interested in performance differences between the different strengths. Figure 8.2 depicts the curve strengths used in the strength test.

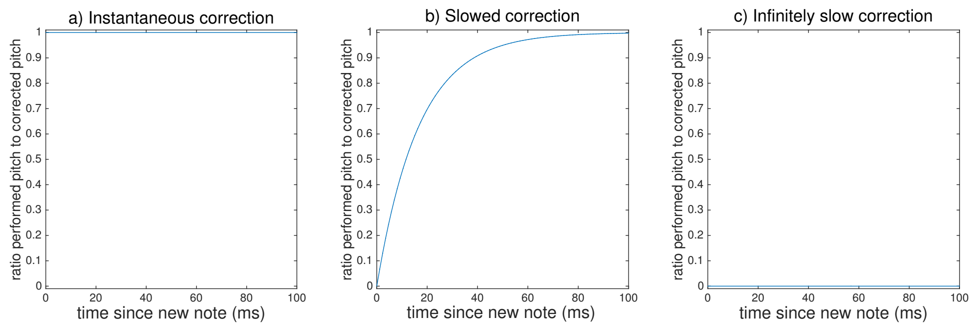


Figure 8.3: Three tested speeds of applying pitch shift. (a) applies snap snap instantly with $\nu = 1$, (b) is a slower transition, gradually phasing in the pitch snap with $\nu = 0.5$ and (c) is $\nu = 0$ which is an infinitely slow transition where the output pitch remains the same as the input pitch.

The third section tested the speed at which the pitch snap is applied. p_t is the target pitch calculated using a strong ($\psi = 1$) snap with an exponential shape[3]. The three speeds tested were with ν set at 1 (a), 0.5 (b), and 0 (c). Again, we chose a control ($\nu = 0$), an extreme ($\nu = 1$), and a middle option ($\nu = 0.5$). With speed we were interested in knowing both whether delaying full pitch correction enabled some level of player correction (question 3, List 2) or if players noticed any impact on sound quality (question 4, List 2). During the first two test sections, we used $\nu = 0.8$, shown in Figure 8.4, as in our pre-trial experiments, we felt an immediate snap with $\nu = 1$ sounded somewhat artificial, but also thought a quick change best so chose something close to, but not $\nu = 1$ as default.

Prior to the test sections, the study included a short two trial familiarization section intended

[3]At full strength, the exponential and linear shapes are effectively the same, so technically the curve could equally be called a linear segmented snap

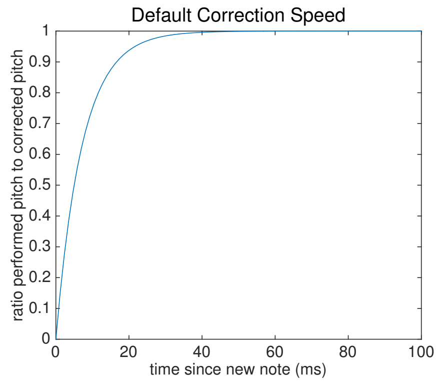


Figure 8.4: Default speed $\nu = 0.8$ used in the expert pilot study sections testing speed and shape. Note that as with other speed depictions, axes are time and ratio performed to corrected. This provides a small amount smoothing when changing notes. After 10ms, the performed to corrected ratio is over 70%, and by 20ms, it is over 93%.

to allow participants to become familiar with the augmented violin and test setup. Each trial required playing an excerpt as presented in the main study that participants were welcome to play as many times as they liked. They were also asked to fill out the Likert questionnaire so that the participants could ask questions prior to the main study. The familiarization section, though documented, did not contribute to results.

8.2.2 Participants

The study with advanced violinists involved 8 participants across the span of two weeks during the summer of 2015 at Queen Mary University of London. Participants, listed in Table 8.3 were recruited through e-mails to two London amateur orchestras, the Queen Mary University of London community, the Guildhall School of Music and Drama Leadership course, and professional violinists personally known to the author. Participants ranged in age from 24-41 with a mean age of 29.5 and a standard deviation of 4.8 years. There were two full-time professional violinists, and one professional violist (who plays violin, though not as the primary instrument), all with 22-28 years of experience (on average 24.3 years) and at least one performance degree. Three participants self-identified as high-level amateur with 15 - 36

Participant ID	Experience Level	Years Experience	Age	Gender
P1	HLA	28	35	F
P2	HLA	36	41	M
P3	HLA	15	30	F
P4 & P8	Professional	23	30	F
P5	Professional	22	26	F
P6	Professional	28	31	F
P7	MLA	5	25	M
P9	MLA	10	24	F
Mean	-	20.86	29.5	-

Table 8.3: Complete expert study user demographics. HLA denotes those who responded they were high level amateurs and MLA denotes medium level amateur. Due to technically failures during her first attempt completing the study, P4/P8 returned to redo the entire study.

years of experience (on average 26.3 years), and two participants self-identified as medium-level amateur with 5-10 years of experience (on average 7.5 years). Though there were clear variations in playing capabilities, no participant seemed to significantly deviate in skill from how they self-identified.

8.2.3 User Response Tasks

List 3 shows the five questions asked at the end of each trial. Participants were asked to rate the first four using a 5 point Likert scale where 5 was *Strongly Agree* and 1 was *Strongly Disagree*. Questions were intended to probe participants' perception of their left hand performance in comparison to the audio output they heard. In order to get participants to think beyond just being in tune and focus more on the relationship of their perceived performance to what they heard, Questions 1 and 2 were intentionally similar but asking using opposing ways of thinking about intonation. Question 3 was targeted to measure the overall perception of the as-heard intonation as easier intonation is both a target and an expected outcome. Question 4 was targeted to capture the ease with which we'd expect people to be able to learn as predictability and repeatability are key components for learning. The last question, which is not a Likert question but an open response, was to allow users to include feedback or experiences they thought relevant.

List 3 Questions asked after every song segment. Questions 1-4 used Likert scale for responses.

1. When I play accurately with the left hand, the sound is in tune.
2. When I play inaccurately with the left hand, the sound is out of tune.
3. I found it easy to play in tune.
4. The pitch output was predictable based on where I put my finger.
5. Comments on pitch experience or other.

At the end of each nine trial section, there were the four questions given in List 4, all open response, along with further opportunity for comments. Unlike the first set of questions, the questions were targeted to capture any aspects unique to a specific session along with general usability and quality of the system. Question 1 was designed to ask if the speed variable was noticeable, and also possibly capture errors in pitch estimation at the beginning of a note. Question 2 was useful for capturing feedback on audio glitches, and judging general acceptability of audio quality, while Question 3 again, focused on more general usability.

List 4 End of Section Questions

1. Did you notice any distinguishable change in pitch after placing your finger on the string? (e.g. glissando or warble effects?)
2. Did you notice any unexpected/unpleasant sound artifacts besides a possible change in pitch?
3. Did you find any pitch modifications from the augmented system distracting?
4. Other comments?

The last set of questions, List 5, at the end of the user's session, were about the feel of the augmentations. These questions were not trying to test any hidden variable, just obtain feedback whether, and to what extent the sensors interfered with regular play. We also asked each player for suggested improvements.

List 5 Final questions on physical sensor augmentations.

1. How did you find the bow balance of the augmented bow?
2. Did the cabling and sensors change how you played?
3. Were there any bow strokes/actions where the augmented bow was an unusual challenge?
4. Other comments [about the augmented bow]?
5. How noticeable was the fingerboard sensor?
6. Did the fingerboard sensor change how you played?
7. Were there any left handed techniques that were a particular challenge with the sensor in place?
8. How would you like to see the augmented violin improve?

8.2.4 Data Collection

Direct audio from the polyphonic bridge was recorded for all sessions, along with two timestamped data streams. The first stream, produced by the LLAV VST plugin was logged every 512 sample window (94Hz at 48kHz) and consisted of sensor data pertaining to left hand finger placement, each string's momentary volume (RMS), estimated momentary performance pitch, and readings from the four sensors on the augmented bow. The second stream, produced by the retuning VST plugin and logged for each hop of 128 samples (375Hz), included the pitch estimate received by the plugin, the target pitch heard by the user, and information on what musical excerpt and test settings were being used. Each plugin's data stream was timestamped in two ways, the first being the current system audio sample count, and the second, a time-since-start generated by the LLAV VST and passed in the data packet to the retuning VST. These two time stamps ensured synchronization between the two plugin's data streams and recorded audio.

The study used two computers, one connected to the augmented violin operated by the author, and one used by participants to complete the questionnaire. This arrangement let the author oversee and control all the software and hardware involved in the study. The author's computer was connected to the augmented violin sensor systems as well as an external multichannel I/O to receive the four separate audio channels for each string from the polyphonic bridge. Reaper was used to host test sessions as Reaper's routing flexibility allowed routing according to Figure 7.2.

The retuning VST was also extended to directly support the study, ordering study trials as discussed in Section 8.2.1 and adding a simple UI shown in Figure 8.5 for stepping through study stages, telling the author and user what music segment was selected for each trial, and letting the author start and stop data recording for each trial.

The participant computer hosted the questionnaire and was arranged, for privacy reasons, to be not easily viewed by the administrator. The survey was done using online Google Forms with the two systems manually kept in sync.

Direction from the author during the study was limited to answering questions, making sure the song segment being played was correct, and that the questionnaire was completed for each segment. A video camera was used to document all sessions and capture record player fingerings, bowing, physical reaction, and interactions between the user and the author.

8.2.5 Measurement and Analysis

Quantitative results are split between the Likert-scale responses provided by users and data from sensor logs on players intonation performance. Results for pitch were calculated based on the linearized pitch estimates of pitch performed and pitch heard collected from the retuning VST logs. In Section 5.4 we demonstrated that for monophonic pitch estimates, our method compared equally well with established state-of-the-art algorithms. Although we did not repeat a full mathematical performance analysis as in Chapter 5, we did verify that estimates using the Aubio Yin Fft pitch estimator [22] with a 2048 sample window were in line with our estimates. We also removed from evaluation calculations any pitch estimates that were clearly incorrect such as those above or below the possible played range (0.45% of samples during play).

Pitch error was determined as the estimated difference in pitch from the nearest chromatic pitch. For example, setting 440Hz A as zero and every integer as a semitone difference so that three corresponds to the 523Hz C above the 440Hz A, a linearized pitch[4] estimate of 3.22 would be considered an error of 22 cents above the nearest note in the chromatic scale,

[4]This uses Equation 7.1 for converting from Hz to hundreds of cents.

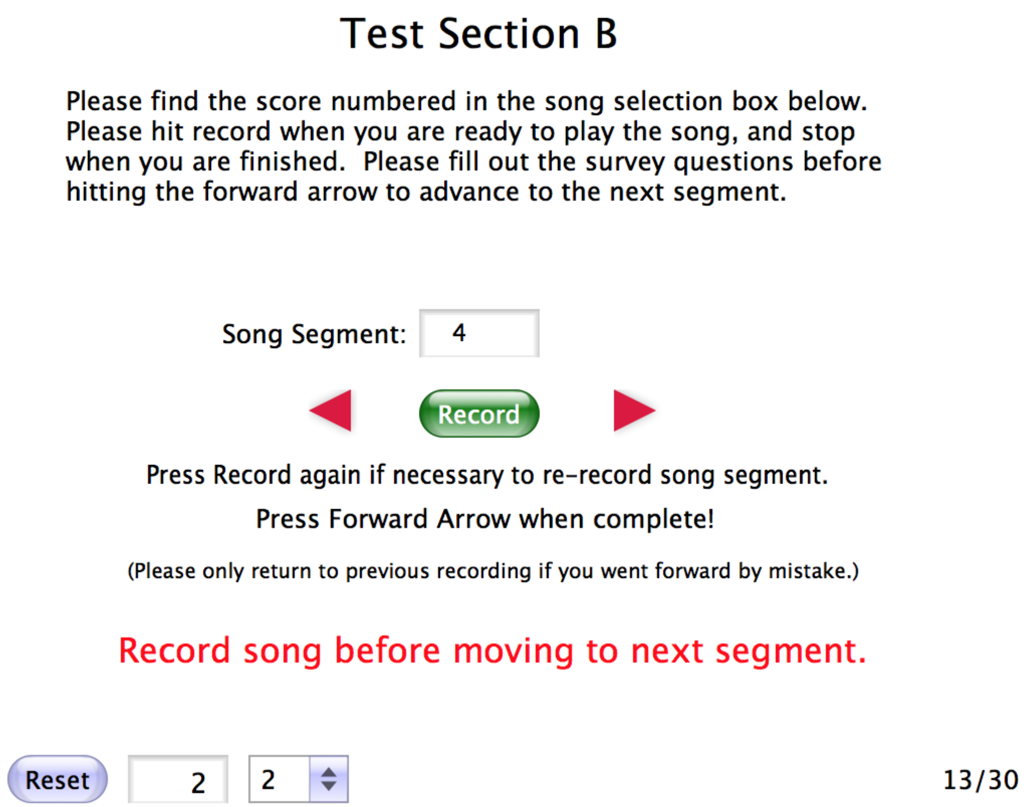


Figure 8.5: The test UI used for navigating through trials. The UI gives information on where in the test the user is, which musical excerpt to play, and requires the administrator to manually start and stop recording the specific trial. Stepping to the next trial automatically updates the excerpt, and the snap settings. Participant number is in the bottom left.

three (the non-linearized pitch estimate of 3.22 is 530Hz). Similarly, an estimated linearized pitch of 2.89 would is labeled an error of 11 cents below the nearest chromatic scalar note, three.

In order to eliminate pitch estimates made when the performer was not playing or when there was insufficient audio amplitude to correctly estimate pitch, the volume for the window (calculated as RMS for the whole window) was computed for each hop. The audio was synchronized to the logged data streams through the time in audio samples. Only hops where the volume was above a threshold were included in the study's data analysis.

There were nine trials where the retuning VST logs used for pitch results were missing. In this case, missing data was reconstructed using the LLAV VST's playback feature (Section 7.1.5) with the sensor data originally logged by the LLAV VST in conjunction with the recorded audio.

Dealing with vibrato

A complicating issue in calculating intonation error was vibrato. Vibrato is an intentional oscillation around a scalar note and was allowed in the study. Typical vibrato deviates $\pm 13-18$ cents around the central mean [179, 54] and will appear as pitch error when in fact intentional. Virtually all participants used vibrato, some people consistently, so removing any section with apparent vibrato from calculations was not appropriate. However, while vibrato does obscure intonation, it is centered around a stable core which is heard as the primary pitch and used by the performer or listener to decide if the note is correctly tuned [18, 54]. The pitch center for the vibrato can be used to estimate whether it is in tune or not.

Vibrato was identified using current state-of-the-art vibrato detection by Luwei Yang [181]. Yang developed AVA, an interactive vibrato and portamento detection and analysis system for MATLAB, that provides interactive and intuitive visualizations of detected vibratos and portamenti and their properties. We used the AVA engine which uses a filter diagonalization method along with Bayes' rule (FDM + BR) to identify occurrences of vibrato. Yang's FDM + BR techniques have an F-measure 0.84 when using frame level assessment, i.e. whether there is vibrato at any given specific moment [180], meaning some vibrato will fail to be detected correctly.

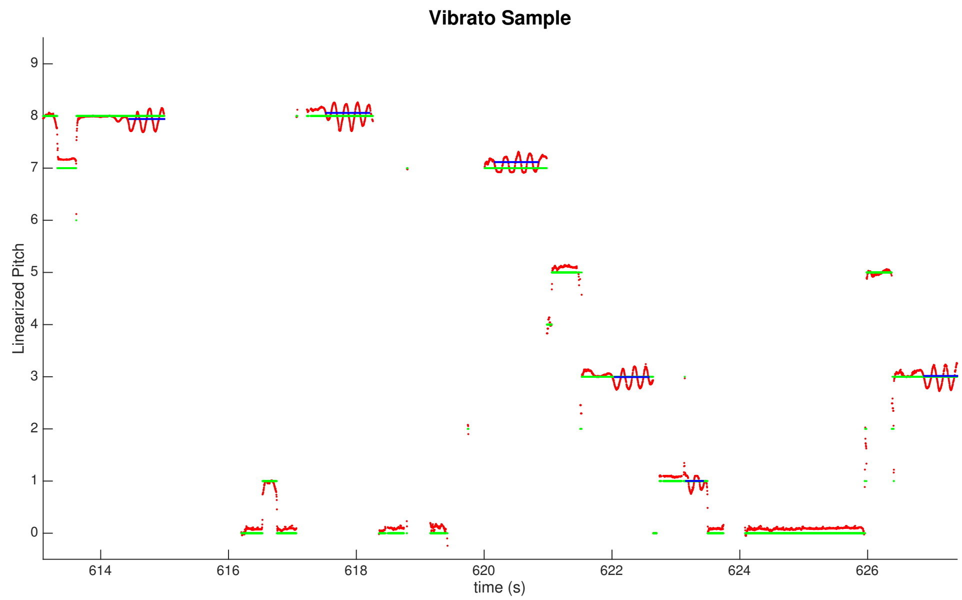


Figure 8.6: Visualization of P5 playing a segment of Schubert's Serenade. The segment contains numerous examples of vibrato, including vibrato starting part way through the note. Red denotes pitch as played by P5 with green denoting the nearest chromatic note. Blue lines indicate where vibrato has been detected and the adjusted as-heard pitch.

Once periods of vibrato were identified, the center pitch was taken by finding the extrema nearest each end and taking the average between them. The heard pitch was then estimated by adding four cents to this center pitch. While there are debates within music pedagogy about whether the intonation center for vibrato is based on the highest pitch performed versus the center pitch performed, research by Geringer suggests violinists using vibrato tune the arithmetic mean of vibrato 2-5 cents below what they would if playing without vibrato [54]. Geringer's finding was supported by our samples, evident in cases such as the example in Figure 8.6 where we can see multiple examples of vibrato being added to a fingered note. The adjusted as-heard center is typically better aligned with the non-vibrato pitch than either the mean or the waveform peak.

Execution Issues

A number of issues during the study execution came up that impacted portions of the study. While there was built in data redundancy to ensure operator error during trials would not cause a problem, there were still some serious non-recoverable errors that resulted in the inability to use some data.

Two significant episodes caused the full loss of two sessions. P2 had to be removed from the test group as the fingerboard sensor hardware was not calibrated correctly. The calibration meant the hardware pitch estimate was significantly above the correct pitch, so far off that retuned pitches were typically a third to fourth above what the participant was actually playing. The only other trial completely removed was performed by P4 who returned to complete the study as P8. The first time P4 took part in the study, a software glitch meant that the audio stopped pitch shifting after the introductory section. As all other aspects of the study were working and logged correctly, the problem was not discovered till the completion of the participant's session.

8.3 Quantitative Results

After removing problematic trials, there were 180 total trials, including 63 trials used to calculate results for shape, 59 trials for strength, and 58 trials for speed. Pitch analysis used 2,664,521 valid samples of pitch or 1:58:25 worth of audio: 1,014,818 samples or 45:06 of audio for shape, 864,621 samples or 38:27 of audio for strength, and 785,082 samples or 34:54 of audio testing speed. For each section, a results table is provided with the mean Likert scores for each question in List 3 along with both the absolute performed pitch error and heard pitch error for each setting. Each table also includes the standard deviation of Likert responses and the RMSE of estimated pitches to provide insight into the variability within data. Means are calculated based on the mean of each user's absolute mean pitch error. Additionally, as the Likert responses can be quite divergent, a table of histograms is provided collating the Likert responses for each valid trial. All control cases, *linear* for the section testing shapes, *no snap* for the section testing strength, and *infinite*, for the section testing speed, should have similar results as all share the same response behavior with output pitch matching input pitch.

	Shapes					
	1) Linear		2) Step		3) Exponential	
	Mean	Std	Mean	Std	Mean	Std
Q1) ...the sound is in tune.	**4.62**	0.67	3.62	1.32	3.45	1.19
Q2) ...the sound is out of tune.	4.38	0.97	3.86	0.80	**3.67**	0.80
Q3) ...easy to play in tune.	**4.10**	0.89	3.24	1.51	3.00	1.19
Q4) ...pitch was predictable.	**4.24**	0.62	2.95	1.28	2.88	1.32
	Mean	RMSE	Mean	RMSE	Mean	RMSE
Perf. Pitch Abs. Error (cents)	**13.69**	17.68	17.25	21.11	16.64	20.59
Heard Pitch Abs. Error (cents)	13.69	17.68	**0.17**	1.05	0.21	1.14

Table 8.4: Results for different shapes of pitch curve. The first shape, linear, acts as a control using the input as the output. The pitch scores are the mean absolute error of the performed pitch as compared to the mean absolute error of the heard (corrected) pitch and the accompanying root mean square errors. Full versions of the Likert questions are found in List 3. Curves tested are those in Figure 8.1. Full-snap heard pitch accuracy is non-zero as the speed, ν, was set to 0.8 meaning the very beginning of notes were not fully snapped.

In order to assess whether the differences in the various means from the contrasting pitch curve settings within each section were statistically significant, we calculated the p-values for both Likert question responses and pitch estimates. p-values for Likert responses were generated using the Friedman test for two-way non-parametric data testing for both statistical difference between all three sets of means within a section and also deriving the p-value for the differences between pairs. Treating the Likert responses as non-parametric raised the differences in mean required to cross the threshold of statistical significance, $p < 0.05$.

p-values for statistical significance between means of performed and heard pitch estimates were calculated using a two-way ANOVA. In order to avoid false significance due to the exceedingly large and repetitive data set of pitch estimates, we used the average absolute error for each pitch trial, rather than the average absolute error for each pitch sample.

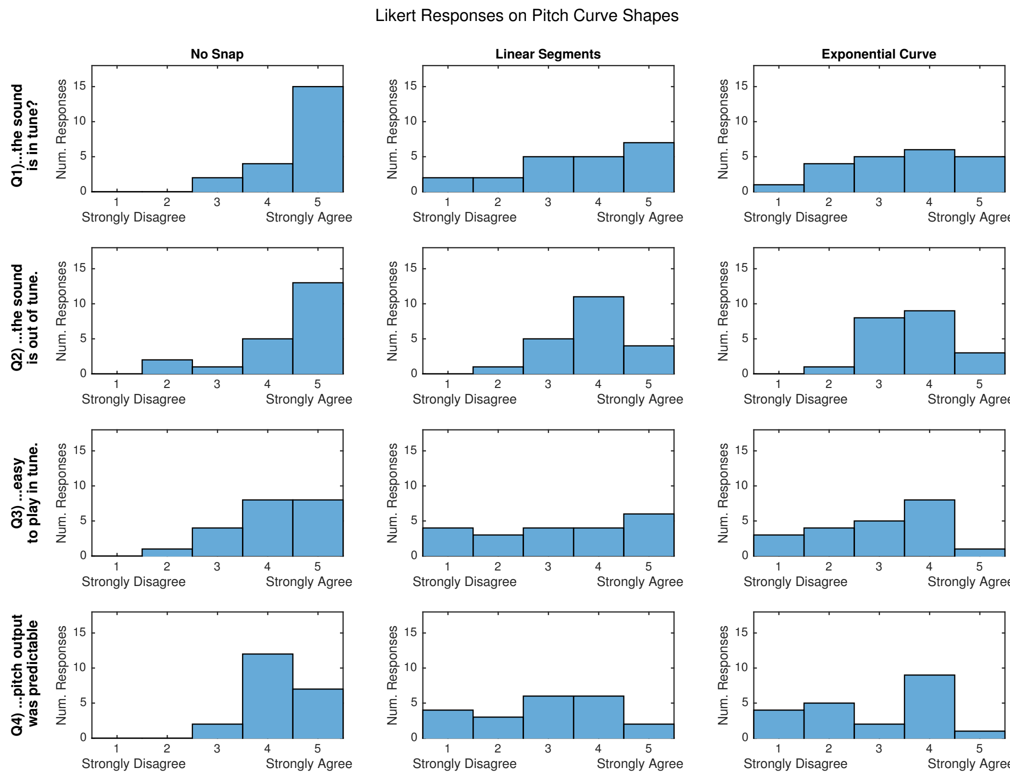


Figure 8.7: Likert responses to questions on curve shapes. The left column of histograms represents answers to the linear response, no snap control case, the middle represent the linear segmented shape responses and so on. Similarly, the rows relate to the question on the left. Full versions of the Likert questions are found in List 3. Curves tested are those in Figure 8.1.

Shapes Likert Response p-values						
	Q1) ...the sound is in tune.			Q2) ...the sound is out of tune.		
	Linear	Step	Exp	Linear	Step	Exp
Linear	-	0.15	0.11	-	0.42	0.31
Step	0.15	-	0.99	0.42	-	0.98
Exp.	0.11	0.99	-	0.31	0.98	-
	Q3) ...easy to play in tune.			Q4) ...pitch was predictable.		
	Linear	Step	Exp	Linear	Step	Exp
Linear	-	0.44	0.18	-	**0.04**	0.06
Step	0.44	-	0.85	**0.04**	-	0.99
Exp.	0.18	0.85	-	0.06	0.99	-

Table 8.5: p-values for Likert question responses and pairs of settings derived using a Friedman test for two-way non-parametric data. Numbers in bold meet a standard threshold for significance of $p < 0.05$. Full versions of the Likert questions are found in List 3. Curves tested are those in Figure 8.1.

Performed Pitch Shape p-values			
	Linear	Step	Exp
Linear	-	**0.001**	**0.007**
Step	**0.001**	-	0.79
Exp.	**0.007**	0.79	-

Table 8.6: p-values for variation in pitch accuracy due to shape. Numbers in bold meet a standard threshold for significance of $p < 0.05$. Curves tested are those in Figure 8.1.

8.3.1 Shape

As shown in Table 8.4, the average response of all four Likert questions was higher for the linear response, the control where output pitch directly matches input pitch. Participants on average more strongly agreed that when they played in tune, the sound was in tune, if they were out of tune, the sound was out of tune, that it was easier to play in tune, and the heard pitch was more predictable. The largest difference in mean was with respect to predictability, with a difference of 1.36 between the control and the exponential curve. p-values, given in Table 8.5, suggest that only in Q4, about pitch predictability, did any differences in preference cross the threshold for significance. The only case with a low enough p-value to suggest statistical significance was the difference in predictability between no snap versus linear segmented step. Figure 8.7 shows that the linear control response received only three scores below the neutral response, while the scores for the other two shapes were widely distributed.

Though the linear segmented step change received slightly more positive mean ratings and slightly higher mean performed pitch error than the exponential response, with a lowest p-value of $p = 0.85$, there was no statistically significant difference between the step and the exponential shapes tested in this study.

In contrast, both pitch snapped cases yielded worse as-played pitch performance than the linear response, no snap control case. In comparison to the no snap case, the average absolute error was 2.95 cents worse when using the exponential snap and 3.56 cents worse when using the linear segmented snap. p-values, given in Table 8.6, both below 0.02 suggest that there is a significant effect on actual as-played performance when using either pitch snap as compared to the control.

Corrected as-heard pitches were significantly more in tune with an average absolute error well below the pitch differential that can be heard by a human. As discussed in Section 2.4, a well trained musician can differentiate between pitches roughly to 2-3 cents apart while non-musicians are closer to 15 cents [117]. The non-zero absolute error is due to the speed at which the pitch snap is applied as $\nu = 0.8$, so there was error allowed at the very beginning of each new note.

	Strength					
	1) No snap		2) Half snap		3) Strong snap	
	Mean	Std	Mean	Std	Mean	Std
Q1) ...the sound is in tune.	**4.48**	.69	4.33	0.73	3.05	1.15
Q2) ...the sound is out of tune.	4.26	.91	3.81	1.11	**3.52**	1.14
Q3) ...easy to play in tune.	**3.76**	1.33	3.67	1.22	2.81	1.45
Q4) ...pitch was predictable.	**3.90**	1.28	3.62	1.05	2.81	1.41
	Mean	RMSE	Mean	RMSE	Mean	RMSE
Perf. Pitch Abs. Error (cents)	**14.53**	18.34	16.27	20.22	17.99	22.13
Heard Pitch Abs. Error (cents)	14.53	18.34	5.89	8.37	**0.22**	1.19

Table 8.7: Results for different strengths of pitch curve. The first strength, no snap, is the linear response control using the input as the output. The pitch scores are the performed pitch absolute error as compared to the heard (corrected) pitch absolute error mean and root mean square errors for each. Full versions of the Likert questions are found in List 3. Curves tested are those in Figure 8.2. Full-snap heard pitch accuracy is non-zero as the speed, ν, was set to 0.8 meaning the very beginning of notes were not fully snapped.

8.3.2 Strength

For the strength of pitch snap, with results collated in Table 8.7, the average response to all four Likert questions was higher for the control where output pitch directly matches input pitch than for either pitch snapping case. Participants on average more strongly agreed that when they played in tune, the sound was in tune, if they were out of tune, the sound was out of tune, that it was easier to play in tune, and the heard pitch was more predictable. However, the half snap was rated fairly closely to the no snap control. These results suggest there is no clear statistical difference in the Likert responses between the control and half snap. Figure 8.8 gives the histogram set containing Likert responses for different curve strengths.

The full strength snap is rated significantly lower than both the no snap and the half snap cases though not consistently enough to always be found significantly different. The p-value between no snap and full snap for Q1, $p = 0.02$, and Q4, $p = 0.03$ suggests a significance difference. It is less likely there is an experiential difference between half snap and full snap

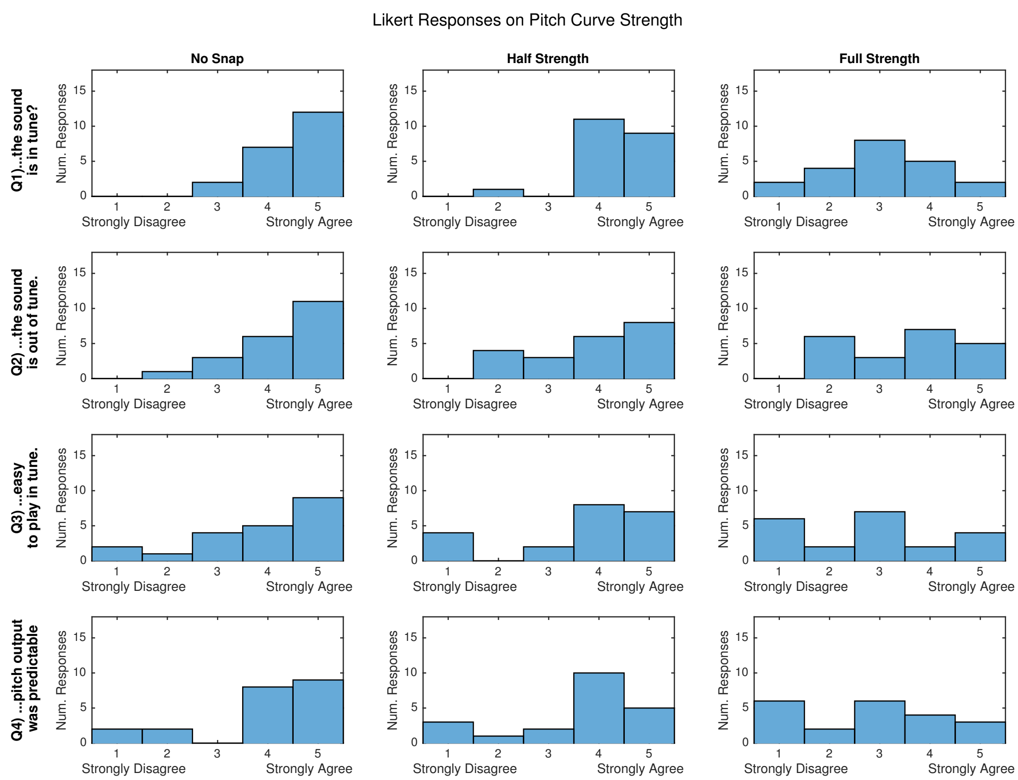


Figure 8.8: Likert responses to questions on snap strength. The left column of histograms represents answers to the no snap control case, the middle represent responses to half strength and so on. Similarly, the rows relate to the question on the left. Full versions of the Likert questions are found in List 3. Curves tested are those in Figure 8.2.

Strength p-values						
	Q1) ...the sound is in tune.			Q2) ...the sound is out of tune.		
	No Snap	Half Snap	Full Snap	No Snap	Half Snap	Full Snap
No Snap	-	0.97	**0.02**	-	0.57	0.23
Half Snap	0.97	-	**0.05**	0.57	-	0.80
Full Snap	**0.02**	**0.05**	-	0.23	0.80	-
	Q3) ...easy to play in tune.			Q4) ...pitch was predictable.		
	No Snap	Half Snap	Full Snap	No Snap	Half Snap	Full Snap
No Snap	-	0.95	0.09	-	0.58	**0.03**
Half Snap	0.95	-	0.17	0.58	-	0.28
Full Snap	0.09	0.17	-	**0.03**	0.28	-

Table 8.8: p-values for Likert question responses and pairs of settings. Numbers in bold meet a standard threshold for significance of $p < 0.05$. Full versions of the Likert questions are found in List 3. Curves tested are those in Figure 8.2.

as only Q1, $p = 0.05$, passes the threshold for significance.

Again, pitch snapped cases yielded worse as-played pitch performance than when there was no snap and the performer could hear what they played. In these trials, full strong snap was 3.46 cents worse than no snap, but half snap was only 1.74 cents worse than no snap, roughly half the difference between no snap and full snap. p-values, given in Table 8.9, testing the difference between different settings were all below 0.05 suggesting that snap strength has a significant effect on actual as-played performance.

Corrected as-heard pitches for the full snap were again almost perfectly in tune with average absolute error of only 0.22 cents. For the half snap, there is a considerable improvement in heard intonation with a 2.76 times reduction in error than as played and a 2.47 times improvement in heard intonation over the no snap control case. The heard absolute mean error of 5.89 cents falls into an error range theoretically noticeable to musicians, but not non-musicians [117].

Performed Pitch Strength p-values			
	No Snap	Half Snap	Full Snap
No Snap	-	**0.005**	$<$ **0.001**
Half Snap	**0.005**	-	**0.005**
Full Snap	$<$ **0.001**	**0.005**	-

Table 8.9: p-values for variation in pitch accuracy due to pitch snap strength. Numbers in bold meet a standard threshold for significance of $p < 0.05$. Curves tested are those in Figure 8.2.

8.3.3 Speed

In the trials for speed, summarized in Table 8.10, the third case, infinitely slow snap, is supposed to match the no snap linear response control where output pitch matches input pitch. Again, the average response of all four Likert questions was higher for the control than for either pitch snapped case. Participants on average more strongly agreed that when they played in tune, the sound was in tune, if they were out of tune, the sound was out of tune, that it was easier to play in tune, and the pitch heard was more predictable. Although participants rated the control higher, findings in the case of speed were not strongly differentiated. The minimum p-value, taken from Table 8.11, between any set of means, (Q4, between moderate and infinite) was only $p = 0.42$.

While there is no clear sign of statistical difference between the three means, responses to Q4 suggest that speed may have an effect on predictability as the combined Friedman p-value between all three settings was 0.06, though as it includes two hypothesis, the threshold for significance drops to $p < 0.025$. The histograms in Figure 8.8 show the diverse set of responses for speed.

Once again, pitch snapped cases yielded worse as-heard pitch performance than when there was no snap though this time differences were far smaller. The largest difference in mean, between moderate speed snap and infinitely slow snap was only 1.70 cents. Trials using immediate snap were only marginally better, 1.44 cents worse than the infinitely slow snap. Table 8.9 shows no p-value near the threshold for significance for differences between means.

	Speed					
	1) Instant		2) Moderate		**3) Infinite**	
	Mean	Std	Mean	Std	Mean	Std
Q1) ...the sound is in tune.	3.62	1.3	3.69	1.16	**3.95**	0.85
Q2) ...the sound is out of tune.	**3.86**	0.85	3.90	0.89	4.05	0.91
Q3) ...easy to play in tune.	3.14	1.54	3.52	1.19	**3.81**	1.11
Q4) ...pitch was predictable.	3.29	1.45	3.33	1.09	**4.00**	1.05
	Mean	RMSE	Mean	RMSE	Mean	RMSE
Perf. Pitch Abs. Error (cents)	17.12	21.23	16.99	21.16	**15.39**	19.24
Heard Pitch Abs. Error (cents)	**0.00**	0.00	0.72	3.16	10.70	16.47

Table 8.10: Results for different speeds of pitch curve. The last speed, infinite, uses the input as the output. The pitch scores are the performed pitch absolute error as compared to the heard (corrected) pitch absolute error mean and root mean square error. Full versions of the Likert questions are found in List 3. Curves tested are those in Figure 8.3.

Speed p-values						
	Q1) ...the sound is in tune.			Q2) ...the sound is out of tune.		
	Immediate	Moderate	Infinite	Immediate	Moderate	Infinite
Immediate	-	0.96	0.89	-	0.99	0.77
Moderate	0.96	-	0.98	0.99	-	0.68
Infinite	0.89	0.98	-	0.77	0.68	-
	Q3) ...easy to play in tune.			Q4) ...pitch was predictable.		
	Immediate	Moderate	Infinite	Immediate	Moderate	Infinite
Immediate	-	0.91	0.58	-	0.99	0.51
Moderate	0.91	-	0.83	0.99	-	0.42
Infinite	0.58	0.83	-	0.51	0.42	-

Table 8.11: p-values for Likert question responses and pairs of settings. Full versions of the Likert questions are found in List 3. Curves tested are those in Figure 8.3.

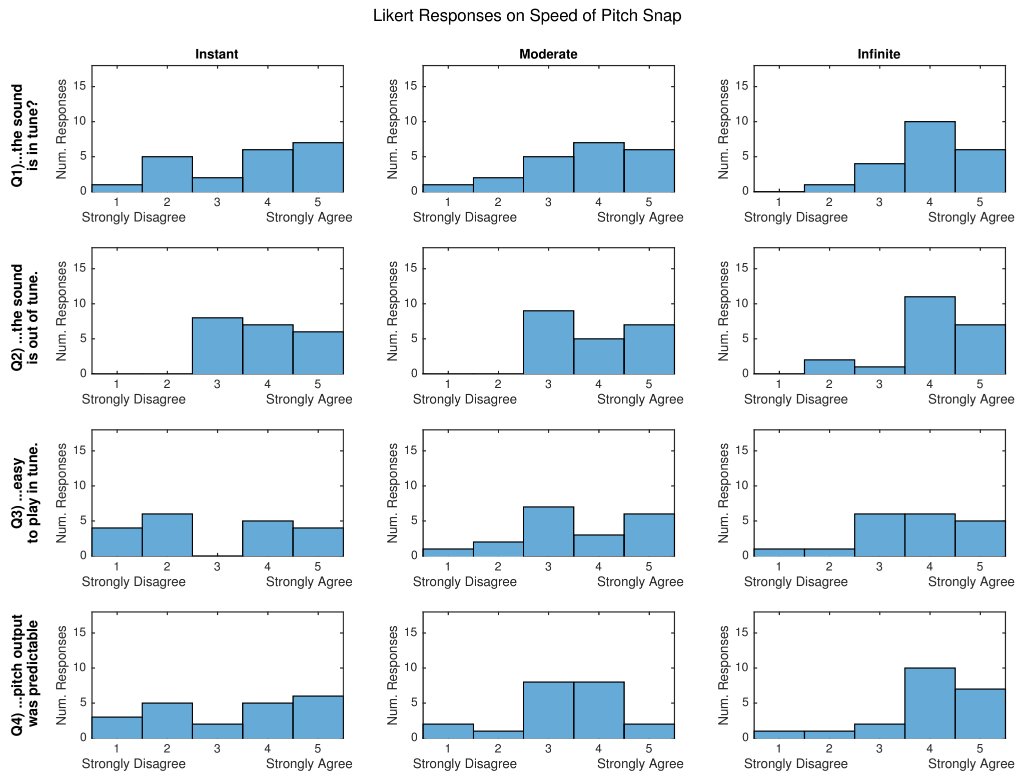


Figure 8.9: Likert responses to questions on speed, or the rate of application of pitch snap. The left column of histograms represents answers to the immediate snap case, the middle represents a slower transition, and the right, where the pitch never snaps. Similarly, the rows relate to the question on the left. Full versions of the Likert questions are found in List 3. Curves tested are those in Figure 8.3.

Performed Pitch Speed p-values			
	Immediate	Moderate	Infinite
Immediate	-	0.95	0.23
Moderate	0.95	-	0.13
Infinite	0.23	0.13	-

Table 8.12: p-values for variation in pitch accuracy due to snap speed. Curves tested are those in Figure 8.3.

8.4 Discussion (with Qualitative Feedback)

We start with discussion of the reaction to the different forms of automated pitch correction prior to discussing augmented violin playability. The overall result consistent across all three parameters, shape, strength, and speed, is that experienced violinists perform better when allowed normal unaltered aural feedback. Tests of both shape and strength showed significant differences between snapped and unsnapped pitch as-played confirming our hypothesis that even without significant shifting, limiting aural feedback by removing error around a performed note negatively impacts pitch performance.

Similarly, though results were not statistically robust, participant ratings of different snap cases were that they consistently considered no snap control cases easier, more predictable, and better in tune. As highly trained violinists would be presumed to have deeply established expectations of the relationship between fingering action and heard intonation we did expect snapped cases to be confusing and that loss of full pitch control might make a player uncomfortable, however we did not expect participants' reaction that corrected audio was less in tune than unsnapped audio. It is clear to see in results that pitch snapped audio is always more in tune than as-played, yet in nearly every case, shape, strength, and speed, the option that did not alter the input output/relationship received the highest average Likert ratings for all four of the questions in List 3.

Before discussing potential reasons for the discrepancy between participants' ratings of as-heard intonation and the numerical results, we first look at the results for each section individually.

8.4.1 Shape

While there was both a clear preference by experienced players for the non-pitch shifted experience along with significantly better as-played intonation, there was very little to differentiate between the two shapes of pitch curve. The minimum p-value differentiating the Likert responses between the two was $p = 0.85$, strongly suggesting that any ratings differences are random, not intentional. Similarly, with $p = 0.79$, there is no clear difference in as-played performance.

In retrospect, this is not a surprising result. By choosing to test the shape using a strong curve, the two shapes end up being mathematically very similar. For instance, with input 49 cents high, the output is zero cents high using the linear segmented shape, and two cents high using the exponential curve shape. 49 cents error is one of the most extreme possible errors, but two cent difference in output is below the ability of humans to differentiate [117] meaning the two snaps will effectively sound the same. Using a mid-strength curve such as in Figure 2.10 would have been more differentiable even if they were less drastically different from the no snap linear control.

Without conclusive results, we are left referring to earlier research on targeting such as [10] and [137] and the assumption that a smoother transition will sound better than a non-continuous function.

8.4.2 Strength

The results for strength suggested statistical difference in all three snap cases. Not only did we find that removing the ability to hear error resulted in poor as-played intonation, but we validated our hypothesis that restoring some level of heard error significantly improves performance, even if still significantly worse than when players hear full error. Of noticeable interest is that the half-snap provided improvement in as-heard pitch without significant cost in terms of user experience. Likert responses for the half snap were only marginally lower than the no snap case, with p-values strongly suggesting no significant differences between the two sets of values. Further, while as-heard pitch using the half-snap had 2.8 times less error than as played and 2.5 times less error than the control, it resulted in half the additional as-played pitch error in comparison to the control than the full snap case. Looking at individuals responses, four users consistently rated half snap as high or even higher than the no snap case.

The statistically significant differences in both as-played and as-heard pitch confirms that half snap allows audible difference from both the full snap in-tune pitch and the no snap case. Players could hear vibrato although it is significantly reduced in width than non-snapped vibrato. Allowing some level of deviation enables the user to retain the feel of control and the physical-aural link that moving the finger one way or another will alter the pitch accordingly. Interestingly, in our study, error in average as-heard pitch for the half snap case (5.89 cents) is above the threshold for pitch differentiation in trained musicians but below the threshold for non-musicians (2.25 cents vs. 14.82 cents [117]). Results for the half snap are particularly encouraging implying we may be able to improve intonation without hampering the performance experience.

8.4.3 Speed

No clear statistical difference in as-played performance between an immediate snap and a moderate speed snap supports the idea that a small change in snap speed does not provide a player enough information to correct pitch. Otherwise, speed results for specific preference between snap speed were inconclusive with no clear preference between immediate snap and moderate speed snap. Questions in List 4 were more appropriate for collecting feedback on the audio quality of pitch snap than the Likert questions in List 3 as the Likert questions were more linked to the overall impression of intonation than catching minor audio effects at the beginning of a note. Not only were differences between Likert ratings for immediate snap and moderate snap negligible but comments collected during the study do not reflect any difference either. No responses to List 4 made clear reference to sound quality issues with the immediate snap case suggesting it may not sound artificial. The one speed related behavior participants did note was the result of a programming bug.

After the study completed, we discovered a bug in the software for the infinite slow setting. Although the infinitely slow snap was intended to match the control where the pitch output matched the pitch input, as programmed, the pitch would instead snap to the nearest scalar note after 266 ms. As a result, what was intended as the control, was instead a delayed pitch snap. This was noticeable to participants anytime they played longer notes. During speed trials, P3 remarked out loud, "It's changing my pitch."

Revised Statistical Significance Testing Due to Delayed Snap

With this in mind, it is worth comparing all three settings in the speed section against a true control case. Compared with the control from the strength tests, p-values for as-played pitch performance for both the immediate ($p = .003$) and moderate speed snap ($p = 0.014$) suggests they are both statistically different from the no snap control. As opposed to the comparison with delayed snap in Table 8.12, this finding is consistent with findings in previous sections.

There is conflicting evidence for statistical significance between the delayed snap case and the no snap case. Compared to the control trials from the strength section $p = 0.181$, but compared to the control trials from the shape section $p = 0.009$. It is interesting that despite the 266ms delay, performance does seem worse with the delayed snap. Possibly either players have not finished correcting their initial pitch or they are disagreeing with the snapped pitch and adjusting in reaction to it.

Table 8.13 compares the Likert responses for speed settings against the responses to the control in the strength tests. Though the p-values for differences in mean are lower than when compared against the delayed snap (Table 8.11) there is still no statistical difference between the different cases.

User Preference

Despite the fact that the infinitely slow speed setting would noticeably snap audio during any held note, the setting still received the highest ratings and the lowest as-played error. It is surprising that the response to Q4 about pitch predictability did not suffer and even matched the earlier controls even though on held notes the pitch would unexpectedly change when playing. Again, this re-enforces participants' preference for being able to hear their own

Speed p-values						
	Q1) ...the sound is in tune.			Q2) ...the sound is out of tune.		
	Immediate	Moderate	Infinite (266 ms)	Immediate	Moderate	Infinite (266 ms)
Control Set	0.35	0.59	0.27	0.72	0.70	0.71
	Q3) ...easy to play in tune.			Q4) ...pitch was predictable.		
	Immediate	Moderate	Infinite (266 ms)	Immediate	Moderate	Infinite (266 ms)
Control Set	0.48	0.79	0.97	0.43	0.37	0.99

Table 8.13: p-values for Likert questions comparing responses in speed section to the responses for the no snap linear response control in test section 2 on strength. Due to a bug, what is called the infinite response setting is actually a snap delayed for 266ms after the note start. Full versions of the Likert questions are found in List 3.

performed pitch, even if the actual audio effect was odd. Considering overall response, it is reasonable to expect that some level of delayed onset for the snap is tolerable.

8.4.4 Impact of Experience on Ratings

Evidence for the impact of experience on Likert ratings is that the lowest Likert scores for tuning ease and predictability (List 3 Q3 & Q4) came from one of the two professional violin players, P5. Meanwhile, the two least experienced players, P7 and P9 rated pitch modifying trials stronger than more experienced players. While not a large enough group to be statistically significant, the two less experienced players' Likert ratings for the different pitch corrected trials were on average 1.08 higher than more experienced players. In comparison, P7 and P9's mean Likert rating for the control was only 0.33 higher than more experienced players. Further, the two less experienced players rated pitch corrected trials above the no snap control for 5 of the 12 questions across all three test sections with minimal difference

	Combined Preferences							
	High Experience				Low Experience			
	a) No snap		b) Full Snap		**a) No snap**		b) Full Snap	
	Mean	Std	Mean	Std	Mean	Std	Mean	Std
Q1) ...the sound is in tune?	4.38	0.89	3.10	0.90	4.22	0.68	3.94	1.15
Q2) ...the sound is out of tune?	4.13	0.99	3.52	1.00	4.44	0.63	4.23	0.63
Q3) ...easy to play in tune?	3.67	1.21	2.43	1.21	4.39	0.68	4.17	1.45
Q4) ...pitch was predictable?	3.92	1.13	2.49	1.13	4.39	0.54	4.11	0.68

Table 8.14: Differences in scoring by more advanced (P1, P3, P5, P6, P8) versus less advanced players (P7, P9) for *no snap* and *full snap* modes. *No snap* ratings are taken from the three control settings in each section. *Full snap* mode is the average of the strong exponential curve used in Study Sections 1 and 2, and the immediate snap in Section 3.

in the remaining cases[5]. The difference in improvement indicates the more favorable ratings are unlikely merely due to more generous raters, but rather the reactions to the pitch tuning experience, may be linked with skill.

8.4.5 Trust and User Experience of Intonation

Returning to the earlier observation that participants' ratings of as-heard intonation and whether they found it easy to play in tune disagreed with the our logging of as-heard pitch, we first verified that the system was working as expected. Using the fingering data logged by the LLAV VST in conjunction with the recorded audio, we were able to play back what participants heard and confirm that pitch corrected audio was indeed in tune. We suspect the discrepancy may come from a variety of issues including discomfort within the pitch snapped experience, distrust of the system, and acoustic bleed. P5 and P6, both professionals, provide good context for the ratings:

> "Yes! I think ... I would rather have heard my inaccuracies as inaccuracies, probably because of familiarity with how it usually works on the violin with years of playing." "Anything that wasn't what I am used to was distracting! In terms

[5]The average difference between extreme settings was 0.35 and the maximal difference 0.84 as compared to and average difference of 1.13 and maximum of 1.79 for more experienced players.

> of whenever it came out as what I wasn't expecting, whether it be from [removal of] vibrato or slight pitch distortion etc. " (P6)

> it is disconcerting not being able to adjust my tuning to what I hear because the pitch doesn't seem to correspond to where I put my fingers ... I couldn't adjust my fingers in order to tune the note I was playing, because I couldn't hear what I was actually playing and therefore didn't know where to move my finger to ... I wasn't in control of my own tuning (P5)

Both participants express a strong reaction to pitch correction and refer to it as being distracting or disconcerting because the system does not behave the same as an unaugmented violin. In corrected cases particularly low scores for predictability (List 3 Q4) likely reflect the negative response to snapped cases responding differently than a participants' expectations more than whether the note heard is predictably in or out of tune. Similarly, a player focused on playing in tune may experience the loss of ability to directly control pitch as making it harder to play in tune:

> difficult to tell at times whether the fingers were correctly placed or not. Similar experience to when playing in a gig with no feedback. (P2)

Playing in an amplified concert with poor on-stage monitoring is notorious for causing intonation issues in a performance [163]. The performer (string player, or singer) can not hear themselves for correction.

Further, it is reasonable to conjecture that participants having a negative overall reaction to pitch correction, might rate corrected pitch more poorly due to the negative experience, not necessarily the specific question asked. Even if the notes heard are actually in tune, because it is disruptive and uncomfortable, performers might rate it negatively.

Separate from participants' experience within the study, one confusing aspect of the study itself was the phrasing of List 3 Q2. For the other three Likert questions, a high rating was what a violinist would associate positively: intonation being more in tune, easier, and more predictable. However Q2 was written as the reverse of Q1 so in Q2, if a pitch snapped case is regularly out of tune, it should score low for Q1, but high on Q2 and vice versa. It is not clear whether participants responses to Q2 were based on interpreting the question as written or that a high rating is associated with better intonation. Only one participant's Likert results appeared to potentially follow the question as written; P5, a professional player, generally rated List 3 Q1 high, List 3 Q2 low, but also gestalt questions List 3 Q3 and Q4 very low. All other participants scored List 3 Q1 and Q2 either similarly or differences were inconsistent. Due to the likelihood of confusion, results for List 3 Q2 are difficult to interpret.

Trust

We see clear examples of players not trusting the system:

> Pitch sounded way off what I was playing some of the time ... Some of the higher notes (on the E string) switched to a different pitch ... It changed the pitch compared to what I thought I was playing. (P1)
>
> ...sometimes I even feel that after correction it's still not the right pitch. (P3)
>
> ...excerpts where the pitch seemed to not match where i put my finger. (P6)
>
> Fairly easy to play in tune- but sometimes open strings sounded out of tune. (P1)

There is evidence that if a player heard a note different than what they expected, they suspected the system was at fault, rather than their own intonation being incorrect. While the augmented violin corrects pitch it does not necessarily correct to the desired chromatic note. Unexpected differences in pitch were predominantly linked to performed pitch being far enough out of tune to snap towards an incorrect note. While the system does make errors, when the pitch estimate is wrong, for the polyphonic electric violin used in this study, the estimate is unstable. The instability results in glitchy audio rather than an identifiable out of tune note.

While it is possible for a violinist to simply play the wrong note, one thing to remember is that, as inferred from Galamian's quote "A performer has to constantly adjust her intonation to match her accompanying medium" [52]. For a violinist intonation is relative [91]. Violinists typically learn left hand intervals rather than absolute positions so that if a violinist starts from a sharp (or flat) position, everything they play is likely to remain sharp (or flat) until the violinist plays an open string or a particularly resonant note which exposes the drift [91, p.51]. With pitch correction, drift is hidden from the player with the side effect that if a player has drifted, they are more likely to trigger incorrect notes.

For instance, P1, remarked a number of times that she felt the system was playing the wrong notes or switched inappropriately. Data from P1 showed she regularly drifted slightly sharp. Additionally, for P1, the e-string was also 20-25 cents sharp[6], a difference not audible when snapped. P1's sharpness combined with the sharp e-string to cause the augmented violin to regularly pull her note on the e-string a semitone above her target. Though the augmented violin was behaving and correcting as designed, the player perceived the system as failing.

Two participants commented about the design choice to snap to the nearest chromatic note. Participants suggested possibly to tune only to notes in a piece's key. Snapping to every semitone results in notes that clearly sound wrong and add to the experience of being out of tune where as snapping to a key allows for a wider margin of player error within which the system snaps to the correct note and makes wrong notes more clear. Snapping to a key was suggested as possibly more suitable for beginners.

The Effect of Acoustic Bleed

It is also possible that participants' rating of corrected cases as more out of tune is due to acoustic bleed. Hearing both the as-played and as-snapped cases would likely highlight intonation error. Without a reference for comparison, intonation errors in monophonic sound are less noticeable or even sound in tune, but hearing both the out of tune performed audio and a correct reference pitch will sound dischordant and worse than error in as-played performance on its own. Although headphones were used to block out the electric violin's acoustic sound, it was not always completely effective. Both P3 and P5 remarked on hearing the sound of the actual violin:

> I can hear the sound from the violin and from the headphone are not the same. (P3)

[6] The violin was always tuned prior to each study though tuning is prone to change over time. P1 was the only case where a string was more than 8 cents out of tune.

> I could hear a slightly different pitch simultaneously. I wonder if the violin was out of tune with the machine and so I was hearing the actual pitch and the corrected pitch? (P5)

Further, in discussion P3 commented that hearing both as-played and as-corrected, she was able to tell when she was out of tune and would try to fix her intonation but found it very difficult. What she heard did not behave the way she expected.

Although apart from P5, other users did not remark they could hear both versions of the audio, subtle bleed may have impacted user's perception on heard audio intonation. It is possible acoustic bleed occurred not just due to insufficient headphone volume, but also bone conduction of vibration from the instrument body through the jaw. If participants were hearing both, it would likely result in corrected cases feeling out of tune, rather than in tune.

Interestingly, both participants, P3 and P5, suggested that exaggerating this experience, clearly hearing both the corrected and original sound, might be interesting and possibly more useful than headphone sound only. Hearing both allowed insight into what they were actually playing that was otherwise removed when the corrected volume drowned out the acoustic.

8.4.6 Trends Within The Study

Looking across all three sections of the study, one clear trend is each section was completed faster. Users may have become more familiar with music and testing procedures or were getting tired and trying to complete the test rather than play well. At least three users requested and were given a break due to fatigue.

Likert Trends

There also appears to be either a level of fatigue, or decrease in possible trust of the overall system based on a drop in scores for the control. Excluding List 3 Q2 due to response ambiguity, in the first section only one control case was ever marked below a 3 in response to any question. However, in the second section the control was marked below 3 seven times repeatedly receiving low marks for ease of playing in tune and pitch output being predictable. Possibly as the study went on, participants' ability to trust the violin or distinguish between different cases decreased. For instance, scores in the third section for moderate speed snap are the highest of any full snap case though that may also be due to users getting used to the snap.

Pitch Accuracy Trends

Similarly, average pitch error also got worse as the study progressed, dropping from 13.69 to 14.53, to 15.44 (though the last number had a delayed pitch snap so intonation differences may be due to the late snap instead). Again, this may be due to fatigue and loss of concentration, or loss of trust in the user's innate finger placement.

There was some expectation that if users had several pitch snapped cases in a row, pitch accuracy would noticeably decrease as users lost feel for the tighter coupling between finger placement and heard pitch experienced with the control, however there was little evidence of this happening. Variation between pitch snap cases seemed consistent regardless of the neighboring pitch snap settings.

Effect of Musical Excerpt Choice

Consolidating pitch accuracy using absolute average pitch error for a given musical excerpt across all strong snapping conditions and control conditions, suggests that the musical excerpt does have some effect on pitch accuracy. As depicted in Figure 8.10, there are somewhat similar trends in pitch accuracy between both the strong snap and the no snap control pitch means. Additionally, testing for statistical significance of excerpt by running a one-way ANOVA using the average pitch error per trial on both sets of data gives p-values of $p = 0.002$ and $p = 0.014$ for the strong snap and no snap conditions respectively.

Again, only one excerpt, Excerpt 4, an arrangement of Brahm's Lullaby, required any shifting. There is no evidence that intonation for Excerpt 4 was significantly different, interesting in the context that Chen's study of cello intonation [25] only included shifting.

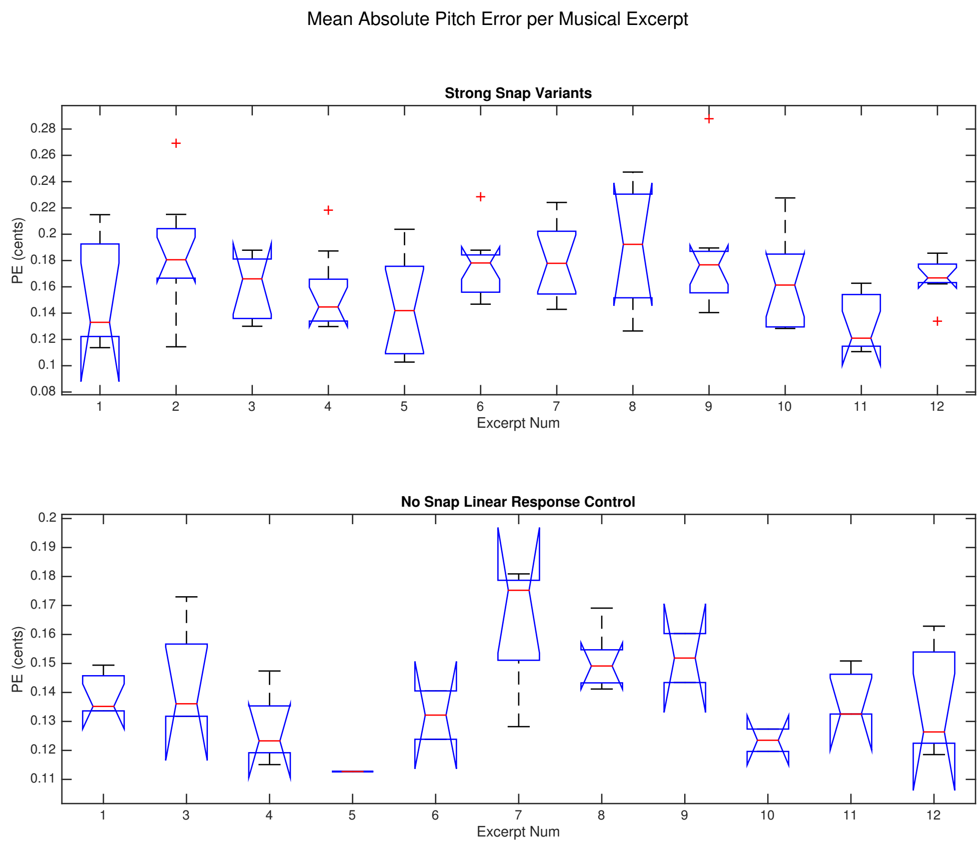


Figure 8.10: The effect of musical excerpt on mean absolute pitch error. Strong pitch snapping conditions were considered as non-control settings in the shape section, the strong snap in the strength section, and the immediate and the moderate snap in the speed section. Control conditions were the no snap linear response control from the shape and strength sections.

8.4.7 Finger Pressure, Audio Glitches, and Vibrato.

Moving on to reactions to augmented violin playability rather than reactions to the different pitch curves, a known prerequisite for our approach to pitch detection is that a reasonable hardware estimate is required in order to reliably estimate pitch and that the hardware estimate requires contact between the finger and the fingerboard. As harmonics are played by lightly contacting the string without pressing the string to the fingerboard, by design our pitch estimation would not work properly with them. Indeed, five participants referenced odd behavior when attempting harmonics. One participant commented directly, "I think there's a problem when I play the harmonics." It also turned out that different players used different levels of finger pressure sometimes resulting in noticeable audio effects, primarily during vibrato.

Depicted in Figure 8.11, loss of the finger position informed hardware estimate results in an audio glitch effect as the pitch estimate jumps around. As non-snapped audio still passes through the pitch shifting process, the glitching distortion will occur both with and without pitch snap. While most participants did not have significant trouble with excessive glitching due to finger pressure, two players remarked on glitches during vibrato as the physical actions in vibrato can lead to varying finger pressure:

> when playing vibrato... or harmonics, some noise occurs. (P7)

> when I didn't use vibrato and concentrated on pitch only, there was no distortion when I hit the note bang on. (P5)

While glitching due to vibrato and harmonics are of less concern as neither are considered beginner techniques and are therefore unlikely to cause problems for our eventual target audience of beginners, two advanced participants found that increasing finger pressure improved audio quality, suggesting their normal finger pressure on the string was insufficient to reliably produce good hardware estimates, "Sometimes the intonation is easier if I give extra proper pressure with my left hand fingers" [P8]. In fact, one of these, an Indian classical player, P6, commented on how the augmented violin made her aware of how light her fingers are on the string as shown in the following series of comments:

> Section 1 - 1: "Vibrato affects pitch, sounds different from the usual way how I

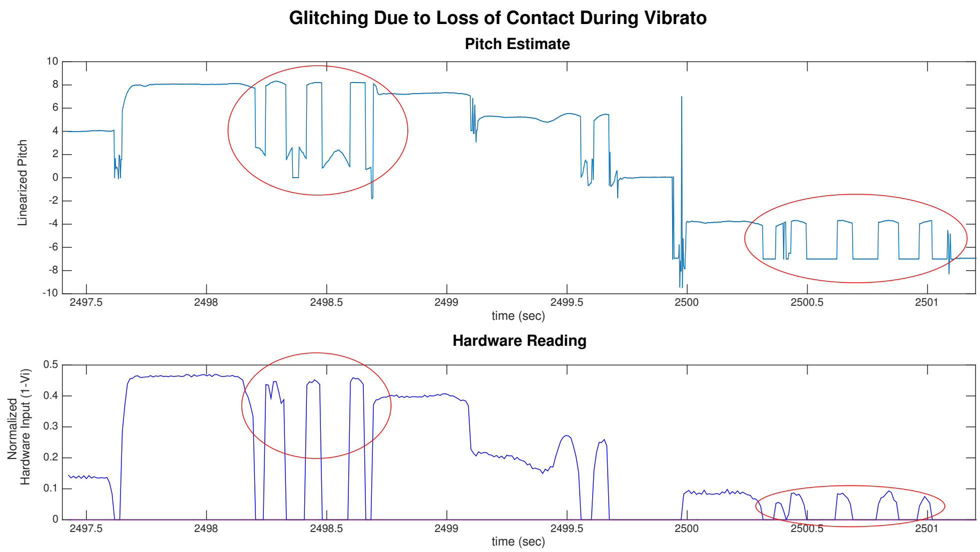


Figure 8.11: Examples of lost finger contact with the sensor during vibrato causing audio glitches. The upper plot is the detected pitch, while the lower is the sensed finger input. The circled areas are where the participant performs vibrato and loses finger contact in the process. These samples are taken from P6.

hear my vibrato"

Section 1 - 5: "Vibrato even more obvious this time in affecting sound"

Section 2 - 3: "Seemed like weight of finger (maybe then of bow) affected pitch slightly"

Section 2 - 5: "Used less vibrato- no idea but maybe this is a sign that I don't put my left hand down strongly enough?!"

Separate from glitches during vibrato, P6 commented "A bit of the wavering of the pitch depending on how strongly my finger was placed or if the left hand was solid." At the end of the session, P6 remarked she intended to practice increased finger pressure the following week as her experience in the study had made her aware of what many in Western traditions would consider a deficiency in technique.

Pitch snap effect on vibrato

While vibrato was sometimes a source for glitches due to pitch estimate error, the absence of vibrato in pitch snapped audio was one of the most noticeable effects of the pitch snapping. With full snap, light to medium vibrato is flattened and removed while wide vibrato, can sometimes hit a neighboring half-step and change the heard note causing an effect similar to a trill (shown in Figure 8.13). As shown in Figure 8.12, reducing the snap strength allows some vibrato effect but at a reduced amplitude. Still the effect was very obvious and every participant commented on the effect of pitch snapping on vibrato:

The system doesn't allow me to do vibrato.(P3)

Vibrato turns into trills. (P5)

The pitch is ok. But sometimes I can not [hear the] vibrato. (P7 after very first trial)

Better do not to use big vibrato, because easy to change for wrong pitch with it. (P8)

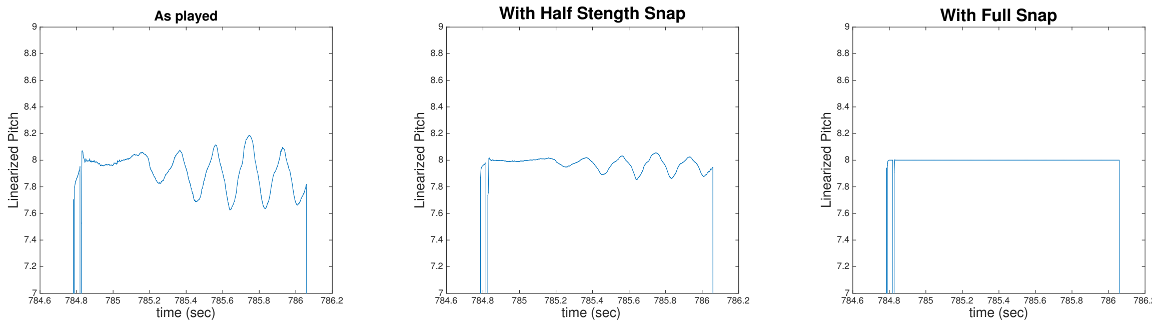


Figure 8.12: Comparison of snapping effects on vibrato. The first shows full vibrato which varies as much as 18 cents above and 36 below the target pitch. With half-snap, $\psi = 0.5$ and $\nu = 0.8$, the vibrato width is much reduced to between 5 cents high and 14 cents below. With full snap, $\psi = 1.0$ and $\nu = 0.8$, only the target note remains.

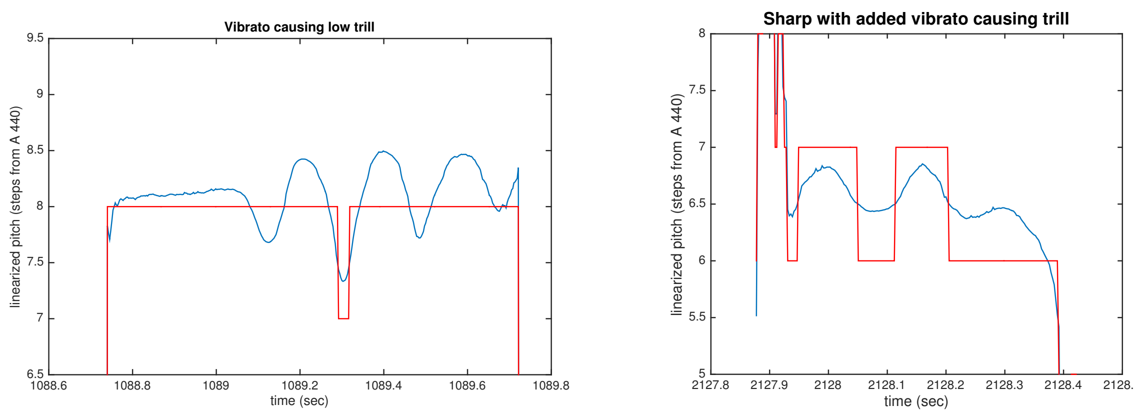


Figure 8.13: Two examples of vibrato being turned into something more similar to a trill. In the first, the vibrato is exceedingly wide and goes more than 50 cents below the intended note, switching where the note is tuned to. In the second, on the right, the core played pitch is sharp so vibrato above the note sometimes goes more than 50 cents above the target, and is snapped upward, creating a small trill effect.

8.4.8 General Audio Quality

We are mostly interested in the augmented violin for practice and learning rather than performance so we were not targeting nor expecting concert quality audio, however we do need audio quality good enough for players to readily accept it. The switch from acoustic to electric violin was in itself a significant change in sound from what participants were used to.

Aside from the glitches during vibrato and harmonics, techniques beyond the level of most beginners, and taking into account that the questions in List 4 specifically request criticism, comments suggest most users found the audio to be acceptable overall. Excluding P1 for whom a subsequently corrected software bug caused clear excessive audio glitching, comments from five of the six remaining participants contain no major complaints about the audio. Four of the six users focused audio quality criticism predominantly on glitches during vibrato and harmonics, with either no complaints beyond those two areas or comments suggesting additional noise existed but was not overly distracting:

> Yes, first, there seems to be some white noises always there...I'm quite getting used to the warbles now. (P3)

> The timbre is acceptable... It is acceptable and OK. The artifacts (are) tiny. Pitch is absolutely ok (if played in a correct technique), although sometimes a tiny noise occurred when I am playing harmonic and other technique. (P7)

> Extra noises are disturbing a bit but are better when I get use to it.... Extra weird noises are a bit disturbing, but manageable. (P8)

> Sometimes it is a little different with the sound of the original sound of the violin. (P9)

A recurring theme in comments is that minor sound glitches occur, but they are easy to get used to. For instance, we asked about audio quality after each section and P9 went from noticing them, responding to List 4 Q2, "Yes, but just a little (distracting), not very important." to later responding "No".

Another positive finding is that only one participant, P5, commented on latency. Although, due to system latency, the latency is much higher (37.6 ms) than the 10ms target, it appears it was not overly distracting. We did not specifically ask participants if they noticed any latency, but expected if it was a problem, participants would comment on it. P5's case was the only time during the study we had major problems with hardware reliability and it affected latency. P5 noticed and commented in cases when latency was increased. The existence of comments on excess latency in P5's case are useful in that they tell us that significant delay did trigger participant reaction whereas under normal working conditions latency was not noticed to be distracting.

The lack of complaint may be partially explained by Mäki-Patola's research. Though 10ms is a generally accepted target latency [49, 72], Mäki-Patola suggests that allowable latency is far more complex. In [105] Mäki-Patola discusses psychological and bio-mechanical aspects of how humans perceive synchronicity, and points out that though pianists can detect a 10 ms latency, most instruments, including the piano have much higher inherent latency. For instance, a soft note on the piano has around a 100ms latency. Mäki-Patola's own research into latency perception in instruments without tactile feedback suggests 30ms is when latency becomes noticeable, but not necessarily distracting [106].

Although we do believe a lower latency would improve the feel of the augmented violin, and could be much improved using a dedicated low-latency audio and sensor system such as Bela [113] in future, the current full latency was demonstrated to be acceptable.

8.4.9 Fingerboard Sensor Useability

One of the explicit goals of the study with experts was to verify that the fingerboard sensor was both effective and non-intrusive. As is clear from Section 8.4.7, the sensor build did not always successfully track vibrato as light finger pressure failed to register, however findings suggest that the sensor was sensitive and accurate enough for typical non-vibrato finger pressure.

The sensor did exceedingly well for feel, receiving no negative comments. The common response to Q5 List 5, *How noticeable was the fingerboard sensor?*, was either "Not at all" (P1 & P5) or similar. Clear evidence of the fingerboard sensor integrating transparently was that a couple of participants did not realize there was anything there until reaching the questions asking about it at the end of the study. P7 commented, "I can hardly feel the fingerboard, very nice!".

8.4.10 Bow Feel and Useability

While the fingerboard sensor was comfortably integrated into the violin performance, augmentations to the bow were more noticeable. Still, most participants found it acceptable though heavy. Half of the study participants, when asked about bow balance remarked, "It is quite heavy, but not too bad to play (P1)," or something very similar. Two participants remarked though that though they at first found the bow "Different (P7)," "It got easier as the test went on but this could be said about any bow (P2)." Two participants (P1, P2) found short cabling between the bow and the wrist made it hard to pay near the frog, an issue fixed when it was re-cabled between P6 and P7's sessions. Three participants (P3, P5, P7) expressed worry that cabling was delicate and they needed to be careful.

The two professionals, P5 and P6, were the only two to suggest they found it harder to accomplish normal bow strokes with P5 saying the bow was not bouncy and that "It was difficult to play off the string." P6's standard playing style involved an unusually high level of bow tilt meaning she had to be continuously careful not to catch the sensors attached to the bow on the string. She was the only participant who repeatedly caught the string with a sensor.

Though the additional weight, cabling issues and clipping of sensors were unquestionably due to the bow augmentations, it is not clear some issues, difficulty playing off the string and overall feel are not also do to the use of a student bow. The difference in feel between a professional quality bow and an inexpensive carbon fibre student bow with the student bow likely far stiffer than participants personal bows. In fact, P9, one of the least experienced players stated "I think the bow is nearly the same with the normal one, I did not find any differences between them. I can play with the augmented bow easily." Factoring in the limitations of the test bow before the augmentation, we consider the overall response to be positive when considered in the context of how sensitive bowing is to small alterations in feel.

Issues with Bow Cabling

One issue participants were not aware of was loss of connectivity between the bow and the sensor processing board. Initially users were required to wear only one elasticated strap which attached to the magnetic connector that took sensor signals to the board. The other magnetic connector dangled from the bow. The dangling connector sometimes caused problems as vigorous playing can pull the connectors enough that the electrical connections are momentarily broken. Pull on the magnetic connection was initially made much worse as the cable between the bow and the connector was slightly short so that there was tension when fully extending the fingers holding the bow.

Additionally, there was a recurrent manufacturing problem with insufficient strain relief leading to broken connections. In minor cases, this may manifest with one sensor freezing but may eventually lead to the whole set of bow sensors being lost. By the time P5 performed the study, the connection between the bow and the magnetic connectors had been pulled enough that the bow had started to fail regularly and the interruptions to the session caused by attempting to clear associated errors clearly negatively impacted P5's overall experience.

The cable length between the bow and the connector was increased after the wiring completely ripped off during P6's test session. Participants were also then required to wear a second elastic strap that attaches to the bow's magnetic connector. The longer cable was more comfortable to play with and, combined with the second strap stabilizing the magnetic connection, eliminated most bow sensing failures.

8.4.11 User Awareness and Preferences

One last effect on results is user awareness and interest into what the augmented system is doing. Participants experimenting or having preferences different from the system design were liable to affect results for any given trial. Many participant comments suggest awareness of the effects of vibrato and left hand finger pressure on sound quality and pitch snapping. Three users made comments to suggest they altered their behavior in response to these effects. For instance, P6 specifically stated, "At some brief moment I intentionally didn't vibrate so I wouldn't have to deal with the uncertainty."

Some users also intentionally experimented with the system response out of curiosity. For instance, P5 asked half way through Section 2, "Do I have to play in tune?" P6 also commented after completing the test that she had intentionally experimented with finger pressure and other system interactions. Participants were free to play how they chose including intentionally playing out of tune, with light fingers, or with heavy fingers.

P6 also pointed out that expectations were potentially overly oriented towards standard Western technique commenting, "Maybe [take] in mind that pitch isn't the only issue and potentially that there are varied techniques across styles and ways of playing that might inform a student, whether beginner or more experienced." Similarly, P5, whom regularly plays baroque and gypsy violin traditions complained, "You can't tune the leading note (i.e. F#) nice and high when the pitch is altered for you to a set pitch".

8.5 Conclusion

We carried out a study asking 8 experts to play simple violin pieces using our electric augmented violin while wearing headphones providing real-time variably pitch corrected versions of their playing. The study consisted of 29 trials separated into four sections, one asking about different shapes of pitch curve, one asking about strength of pitch snap, and one asking about different speeds of pitch snap, and one for familiarization. We successfully achieved our first goal, validating that the augmented violin worked sufficiently for use during most violin performance. Fingerboard sensors did not interfere with play and pitch estimates were sufficiently accurate for use in pitch correction during normal performance without clearly degrading sound. An exception is that pitch tracking does not perform reliably when using vibrato or harmonics. Although we did not assess bow tracking accuracy, we verified that the augmented bow was useable and did not obviously impede normal performance. We found that although our present cabling approach was functional, effective robust cabling that is easy to remove remains a challenge.

Our second goal was evaluating the effects of different forms of pitch correction as a foundational step for studies into complexity management. We had four main target questions for the study which are listed in List 2. We wanted to confirm our belief that aural feedback is essential for effective intonation in string playing, even when performing basic music. We found statistically significant differences between normal play with no pitch modification and the as-played pitch error in pitch corrected cases where the aural feedback for correcting intonation was removed. Pitch corrected performances were worse as played and users also responded negatively to the experience.

Our study of snap strength also affirmatively answered our second question, whether providing participants with partially pitch corrected audio would result in improved as-played pitch over the fully pitch corrected case that concealed error. Particularly promising for complexity management investigations was that restoring limited error through a partial pitch snap, though still causing statistically worse as-played performance than not correcting pitch at all, provided sufficient control that players did not experience it significantly different from normal play. We found that marginally slowing down the pitch snapping effect did not appear to enable players to correct pitch.

We were also interested in learning whether players preferred different shapes of pitch snapping curve and speeds of pitch snap. Our experiment design meant we were unable to make and clear conclusions about preferred shape of the pitch curve. Additionally, study participants did not display any preference between immediate or pitch snap applied at a moderate more moderate. In fact, participants rated an unintentionally included delayed pitch snap version strongest. The ambiguous results suggest that speed of pitch snap is not a major issue, though could possibly benefit from a more focused investigation.

Beyond our primary goals and questions, we found that expert violinists generally disliked the full loss of pitch control despite snapped cases eliminating heard error, though there was evidence that the degree of dislike depended on experience level. Additionally, participants reported that corrected trials were more out of tune than uncorrected trials. Responses may be justified by participants preference for the unsnapped experience, but may also be an effect of acoustic bleed causing them to hear both corrected and uncorrected audio simultaneously.

Chapter 9

Extended Study of Pitch Learning with Beginners

In Chapter 8 we verified the viability of the augmented violin and explored its use to simplify pitch control on the violin. In this chapter we proceed to our second study, focusing on our primary objective, whether we can use technology to aid violin learning. We chose to focus on intonation as it is one of the most glaring tasks impeding musical achievement in early violin learning. We conducted an extended study with beginner violinists to investigate three different real-time technological approaches for assisting pitch learning: one using aural feedback, one using visual feedback, and one using both aural and visual feedback. Additionally, we included follow up to our pilot study, this time looking at beginners' response to automatic pitch correction. The study was conducted through lessons 'in the wild' allowing us to demonstrate the augmented violin outside of a laboratory setting and test learning in a more realistic context.

9.1 Study Motivations

Sitting in on a group class, Kate Conway, head of London's Suzuki Hub asked her students what was the most important issue to focus on while playing a piece. To demonstrate, she played a piece twice. The first time she played with good intonation but poor tone and bow technique. The second time she played with excellent tone but poor intonation. She then asked the 15 or so students which one was worse to listen to. The students quickly agreed the example with poor intonation was clearly less enjoyable.

In the extended study with beginners, we continued to focus on pitch due to its importance in performance. Intonation correction being a constantly iterative process, we wanted to investigate whether real-time feedback could assist a beginner student's intonation.

9.1.1 Aural Guides

The decision to study whether an aural guide can be a useful practice aid was inspired both by suggestions from participants in the pilot study and also the long standing tradition of playing with the teacher. As discussed in Section 8.4.5, two participants commented that hearing both the corrected and uncorrected audio might usefully highlight error. During practice there is no teacher to help identify intonation error. Even when a teacher is present, it is often not practical to regularly interrupt students to fix their intonation. Playing with a student is more practical as it provides an immediate example against which a student can quickly self-assess their performance,

Further, Shinichi Suzuki is quoted as saying [162, p.126], "I have found that young children who have been given a chance to listen to good music acquire a good sense of music." Similarly, Kohut [88, 61] states, "The quality of our musical conception is directly influenced by the quality of the musical performances we hear.... it is therefore critical that the 'musical ear' be programmed with superior musical concepts or images." Not only does an audio guide provide a means to identify error, but it also means students are regularly hearing the piece played in tune. Hearing the piece correctly theoretically strengthens the internal ear, the internal sense of pitch.

When working on correct intonation during practice, a student has four common options: playing with a recording, a piano (or equivalent tuned tool), using a tuner to analyze performed pitch, or generating a continuous drone for comparison[1]. Though each method has its

[1]Hub teacher Kate Conway recommends students use a tuner when appropriate. Tuning against a specific

usefulness, they all have substantial drawbacks.

Playing along to a recording is of limited use due to a recording's fixed tempo and linear progression. Practice should often be carried out well below performance tempo and is often focused on short excerpts. Play with a recording is usually only feasible once a student has learned the piece to a sufficiently high standard. Playing with a piano (or similar) requires both a piano and a second person, or for the student to switch back and forth between two instruments. Interrupting to check with another instrument interferes with practice flow and requires remembering what the target note sounded like once back on the violin.

Many modern digital tuners, available either as specialist accessories, smart phone apps, or computer software, analyze pitch played and display the nearest scale note and how far away the performed pitch is. Practicing using tuner feedback requires watching the tuner for pitch information and adjusting played pitch accordingly. Besides the demand for visual attention, a drawback of this method is that as most tuners are designed for instrument tuning and optimize for accuracy at the cost of speed. Tuners are not designed for fast response making them too slow for use in all but the slowest practice [98].

Lastly, a less common method involves setting a tuner or other signal generator to continually produce a drone so that a reference pitch is immediately available to the ear for comparison when needed. The drawback is that often only one reference note can be played at a time and unless turned off, which requires interrupting play, continues to play even when not wanted.

We are not familiar with any research into the use of real-time aural guides in pitch training. Using the augmented violin and retuning VST we should be able to provide an audio guide for correct intonation. Mimicking the teacher playing with the student in traditional pedagogy, we ask:

does providing the beginner with a simultaneous, pitch-corrected aural guide help improve a student's pitch accuracy?

Use of the augmented violin and the retuning VST will rely to some extent on students playing near the correct note, but in work on intonation, we expect that to be a reasonable

drone was taught to the author by Tufts University and MIT violin teacher Sarita Uranovsky expectation.

9.1.2 Visual Aids

In her discussion of how to design learning aids for instruments, Johnson [76, p.48] argues for interactions to use a different modality than those already used in a task. Based on Multiple Resource Theory and Threaded Cognition, Johnson suggests that aural feedback may overload aural processing as the brain is already using auditory centers for listening to oneself. However, when not needing a score, violin does not require significant visual input meaning the visual mode is free. Hence, along with our initial idea of an aural guide, we were interested in knowing

does providing the beginner with realvisual feedback for correcting pitch help improve a student's pitch accuracy?

As previously discussed, using a tuner as visual feedback is a potential aid for correcting pitch during individual violin practice, however tuners are only considered effective when playing slowly [98]. In Section 3.3.1 we reviewed various technologically assisted tools for pitch training and all were based on forms of visual feedback. Work [100, 98, 175, 35] was developed with either singing or violin in mind and all but one, [35], performed assessment and feedback through a score following approach. Score following is itself a technical challenge, requires a linear fixed interaction, and requires substantial preparation as study pieces must be appropriately prepared for the score following system. Only two of the four tools mentioned are real-time, similar to our target interaction. de Sorbier [35], which created an augmented reality version of the violin with virtual frets, is the only real-time pitch aid we found for the the violin not based on score following but the given accuracy was insufficient for fine tuning.

With our high accuracy low-latency pitch estimate, we are interested in seeing whether we can build a rapid response tuner effective for practical practice. An advantage of providing visual feedback on pitch performed is that it is convenient to display not only whether a player is sharp or flat, but also what note is being played. If a student knows a target note is a $G\sharp$ but is not confident what note they are playing, it is easy to include this information in visual feedback. Even if a student starts out unfamiliar with note names for tones they play, repeatedly seeing a note name associated with a specific fingering, should help students learn them.

9.1.3 Aural and Visual feedback

Another recommendation of Johnson's [76, p.126] is that mixed modality can often be best. Mixed modality can combine to emphasis a learning target or a student can refer to one modality in some cases, and another when that first modality is inappropriate. We already know that when using a score visual feedback might be difficult to use and there may be other cases where it is not appropriate. In such cases, aural feedback may be more useful. Similarly, if there is confusion whether the guide pitch is the correct note, or the student is having trouble deciding whether he/she is sharper or flatter than the guide note, visual feedback may provide better clarification. With these ideas in mind and the availability of both aural and visual feedbacks, we ask:

does combining modalities, using both an aural guide and visual feedback, help improve a student's pitch accuracy?

9.1.4 Further Investigation of Complexity Management

While our study with beginners focused primarily on more traditionally inspired practice tools, we wanted to continue with our work looking at violin simplification. The pilot study with experts presented in Chapter 8 gave us many valuable insights into potential approaches to complexity management, but it did not include any beginners, the audience most likely to benefit from the idea. Further, one of our findings was that response to the pitch snapped violin appeared to be linked to experience. Intermediate players were more positive about the pitch corrected augmented violin than professionals. A study with beginners lets us ask:

how will a beginner react to a violin with simplified intonation?

If, like professionals, beginners strongly dislike the pitch snapped experience, pitch correction will prove unsuitable as a potential simplification.

An additional finding of the pilot study was that experts rated real-time corrected pitch as more out of tune than as-played pitch despite corrected audio being more in tune. Though it is possible experts rated corrected pitch worse than as-played pitch due to discomfort with the experience, one of the other potential causes was audio bleed. Again, for our ideas on pitch simplification to work, we need to better understand the inconsistency in results and ensure that audio bleed (where the acoustic sound of the violin is still heard alongside the pitch-corrected version in the headphones) is not a problem.

9.2 Pre-Study Verification: Follow-On from Pilot Study

Prior to the beginner study we performed some informal intermediate trials to explore issues raised but not resolved in the pilot study. First and most significantly, the beginner study differs from the pilot study in that it uses an acoustic violin instead of an electronic one. Choosing to perform the beginner study with children required finding smaller violins and we discovered that electric violins and the polyphonic bridge are only readily available in full-size. Considering that part of our design objective is that augmentations should be useable with every day instruments, it made sense to switch to an acoustic violin. Additionally, a number of participants in the pilot study had remarked they were put off by the sound of the electric violin and how different it was from their normal violin. Presumably, switching to an acoustic instrument would reduce this reaction and reduce the barriers of acceptance for the augmented violin.

The acoustic augmented violin in Section 5.7.2 was designed for use in the beginner study. We had two major concerns with the acoustic violin: 1) whether our low-latency pitch tracking would tolerate a monophonic input, and 2) whether the acoustic sound would be too loud meaning acoustic bleed would be a major problem. The monophonic acoustic low-latency violin is discussed in Section 5.7.2 but it suffices to say that in informal studies, we found pitch tracking with the acoustic violin was at least equal to the electric violin. Further, any software issues causing significant audio glitching during the pilot study were corrected if not already sufficiently resolved.[2]

[2]Hardware related issues with glitching due to loss of contact during vibrato and harmonics do remain but

9.2.1 Ensuring Pitch Guide Clarity

We performed pre-study informal tests on audio bleed. Part of this was due to the switch to the acoustic violin, but it was also strongly motivated the desire to investigate the inconsistency in the pilot study between the rated intonation of corrected cases versus the in-tune corrected audio produced. We asked an adult beginner to test both the one headphone ear and the double headphone case.

We found that a significant issue with the one ear case was that it could be difficult to distinguish which sound was the player and which sound was corrected. Adding heavy multi-band compression to the violin audio resolved this confusion though it did make the sound more artificial and boosted background noise. Additionally, headphone audio had to be quite loud in order for it to compete with the acoustic sound. As the beginner stated:

> When we first tried with the normal acoustic sound, the sound in my ear was really nice, but it sounded just like the actual violin. So I could hear myself on the violin and I could hear that same sound through the headphone or the speaker. And this was actually really difficult and an extra thing to think about when I was trying to follow the pitch in my headphone. Because it was hard to differentiate. I didn't enjoy the compressed sound as much as it sounded computery and not as nice sound quality, but it meant for what I was trying to do, hear the snapped pitch and be able to adjust on the violin, that was better with the computery sound. And it mean that my brain didn't have to compute which one to follow. It meant it was really obvious and I could get on with adjusting my pitch.

Similarly we found that it was possible to effectively drown out the acoustic sound when using double headphones. Volume had to be loud but remained comfortable. The pre-trial test found that with compression on, corrected audio did indeed sound in tune. We save further in depth discussion of audio bleed for Section 9.8.5.

As the pilot study suggested that speed of snap was not a significant issue, we did not test speed of snap in the beginner study but, not having a clear best setting from the pilot study

are not of significant concern as neither are beginning techniques and our target audience of beginners was not expected to use them.

we tested it in pre-trial tests. We opted for a moderate speed snap with $\nu = 0.5$ for the beginner study as though differences were subtle, the moderate snap was preferable.

9.2.2 Snapping Chromatically Versus To Key

Lastly, based on the suggestions in the pilot study that snapping to a piece's key might be better, we tested snapping chromatically versus diatonically. Using the setup for the aural guide, where the participant wears one headphone, we asked the beginner to play a major scale. When snapping chromatically, the beginner said:

> I was getting it wrong. I didn't know if I was sharp or flat because it was correcting me to the wrong note. And as a beginner, I just didn't know. I knew it was wrong, but I didn't know why. The actual snapped one was wrong so it wasn't helping me to know if I was flat or sharp.

After snapping to key, our beginner reported:

> That was much clearer. There's so much information coming in. I'm listening to the acoustic, listening to the snapped one, seeing the visual, trying to play in tune, and when I had to process whether the snapped one was actually right or not, there was this whole other thing I had to think about before I adjusted, but when it was limited, I knew I could trust my ear and it was straight to adjusting whether I was flat or sharp. So it was one less thing to have to compute in my head.

Because we only had full-sized bows for bow tracking, only participants using the full sized violin used the augmented bow. As with the pilot study, we have not explored the bow data collected in this thesis, limiting feedback requests to usability only. The bow used was the same as in the pilot study.

9.3 Expectations

As it is inspired by teaching tradition, we expect aural feedback to demonstrate strong potential as a learning aid. Though we are providing feedback through a mode already in use, it is clear that musicians are able to listen to and process more than just their own playing; ensemble musicianship is built on listening and responding to other musicians. Pitch corrected aural feedback should integrate well into the intonation process loop from Section 2.4.2, particularly steps 2 and 5 from Figure 2.8(b) as it provides an external version of what should be played to support the player's internal version. If internal audiation is weak, the pitch corrected version fills the void.

Similarly, if properly designed, once beginners understand how to interpret it, we expect beginners' pitch accuracy should benefit from using real-time visual feedback. We expect visual feedback should in some ways be more immediate than aural. By immediate, we are referring to the fact that with visual feedback we can be explicit whether a note is too low or too high and how the student must move the finger to correct. Because aural feedback theoretically pairs well with the intonation process loop, we are assuming beginners should intuitively understand how to identify whether they are too low or too high themselves and react accordingly. However, if that assumption is wrong, even only occasionally, visual feedback may be more clear.

Despite being more immediate, we expect visual feedback to be slower to process than aural feedback. Hanson et al [65] studied reaction times to different modes of interaction finding that people react noticeably faster to auditory stimuli than visual stimuli (161.3ms versus 206.9ms in unimodal trials). Additionally, we expect aural feedback to be more reliably useful. As mentioned, visual feedback requires direct visual attention and takes up a specific physical space. Shifting focus to visual feedback requires a physical action. In comparison, if a clash in pitches catches a beginner's ear, they only need to shift mental focus.

Additionally, a major goal in learning intonation is development of the internal ear. It is common place for beginners to have stickers or thin strips of tape on the fingerboard designating where fingers should go. The markers give students both physical and visual markers to target. However pedagogs warn against over reliance on such tapes [91, p.51], with Martin saying [145], "Dots become a visual crutch that students depend on; they're not listening to themselves, they're just going visually." Our visual feedback potentially suffers the same drawback.

Lastly, we must consider the individual student. Quoting Johnson [76, p.53]:

> One currently controversial (but also popular) theory in education suggests that individual learners have preferences to learn using particular modalities. The theory claims that learners fall into three or four different categories. These are: Visual learners, Aural learners, Kinaesthetic learners and learners who prefer reading/writing [45]. There is still a lot of debate about whether learning styles are a useful concept for teaching and learning [159] or have any physical basis in the brain [53]. However the general idea that people have individual preferences towards using particular modalities should not be overlooked.

Especially if learning preferences exist, we expect some students will find one mode of feedback more helpful than another and that reactions will be diverse.

Every discussion of one form of feedback being more helpful than another explains why we expect combined feedback to be the most helpful. Either form of feedback can be ignored or used when convenient. Overall we hypothesize that combined aural and visual feedback will be the most preferred intonation aid and yield the best performed pitch. We hypothesize that aural feedback will perform better in both preference and pitch perfomance than visual feedback as it is passively available at all times and fits more naturally into pre-existing training. Lastly we hypothesize that some form of feedback will be outperform no feedback though we do recognize there are cases where a student may want to actively disengaged from all pitch feedback assistance due to demands on attention.

For follow-on from the pilot study, we are unsure whether beginners will react positively to the pitch corrected experience. As Johnson points out [76, p.42], "Focusing on carrying out a skill can improve a performance in novices but impair performance in experts" so there is reason to believe beginners will react differently, but negative responses in Chapter 8 were stronger than expected. Still, at a minimum, we expect beginners will respond positively to being able to easily play in tune.

9.4 Extended Study Design and Execution

We conducted lessons with beginner students using an acoustic augmented violin in order to compare the effects of four different pitch feedback methods on pitch execution and perceived helpfulness. Each lesson included study specific repertoire before proceeding to material chosen by the student. At the end of each lesson, we included a section asking students to play while hearing only automatically pitch corrected audio, similar to conditions in Chapter 8. The four methods of pitch feedback are:

1. Aural feedback
2. Aural and visual feedback
3. Visual feedback
4. No feedback

Aural pitch feedback was provided by using the retuning VST from Section 7.2 to automatically snap the student's performance to the nearest allowed pitch in a music piece's key. We asked students to wear one ear of a headphone so that the student could hear the guide pitch in one ear while the other ear was free to listen to their acoustic performance. The arrangement enables a student to compare the guide pitch with what they are playing. During lessons using aural feedback, we used a fully snapped pitch curve, with a snap speed[3] of $\nu = 0.5$.

Visual feedback was provided using the retuning VST. As per Section 7.2.2, the retuning VST was extended to include information on what chromatic note was being played and a colored bar representing whether a student was above or below it. The location, color and size of the error bar represented the direction and level of error. A small green bar indicated as-played pitch was close to the note whereas a tall red bar indicated the as-played pitch was far from the chromatic note.

The combined feedback case used a single headphone ear playing pitch corrected audio and provided students with the visual feedback graphic. Students were allowed to use either form of feedback as they saw fit. The last study case, no feedback, was a control case teaching a

[3]See Section 2.5 and Section 7.2.1 for full details and representations of used pitch curves

lesson as normal but with students playing on the augmented violin. Use of feedback in all lessons was optional. Students were told that they were allowed to remove headphones if they found them distracting or uncomfortable and nor were they required to look continuously at the visual feedback.

As well as the four main lesson feedback methods, we included a separate short reduced feedback section in each lesson where students wore both ears of the headphones. Following on from the pilot study investigations of complexity management, in this section we attempted to block the acoustic sound allowing participants to only hear pitch corrected audio. Pitch was either fully snapped, or half snapped ($\psi = 0.5$) using an exponential curve (Equation 7.4) similar to the strength section of the earlier experiment with experts in Chapter 8. The only difference in settings was that we used the moderate speed snap, $\nu = 0.5$, tested in the third section of the pilot study. Aside from complexity management investigations, including a segment using full headphones allowed us to compare the various feedback options with what is essentially a reduced feedback option that reduces or eliminates normally heard error around a note. Within this chapter, we refer to the reduced feedback option with both ears of the headphone as double headphones (DBL HP) to distinguish it from the aural feedback method which used only a single ear of the headphones also often referred to as single headphone.

9.4.1 An 'In the Wild' Environment

Both our primary research themes, design of the augmented violin, and investigations into violin learning aids, are driven by a desire for practical tools that can eventually be deployed in a home practice session. While in this thesis, we have not yet reached the stage to hand out instruments to take home, we did want to test in a more realistic environment. Hence, this study took place in the context of real-world lessons.

As most active beginner violin students are children and we were asking students to come for four lessons, it was necessary to structure lessons in a way to keep students engaged. This meant that instruction had to hold students' attentions and varied between structured study tasks and student driven tasks. In order to get students returning week after week, it was necessary to ensure material and topics covered were valuable. This necessity resulted in a continuous compromise between running lessons as lessons and running lessons as a study.

The lesson context meant that students would be challenged with complicated learning tasks while at the same time, having to adapt to new feedback tools. The drawback of this is that students are not putting full concentration into best use of the feedback tools, but the benefit is that we are able to see the real-world interactions and the boundaries of when feedback is or is not useful.

Additionally, teaching a lesson is not just mentally demanding for the student, but also the teacher. Running the augmented violin and associated documentation tools was an involved process. As most teachers have a specific set of lesson plans and are paid to teach to a basic standard, it proved difficult to find teachers willing to actively take part in lessons using the augmented violin, either with or without the author assisting. As a result, although it risks introducing bias, being a competent violin teacher, the author taught all lessons herself.

Impacts of an 'In the Wild' Context

Rose Johnson writes extensively about the challenges of studying practice aids 'in the wild' in her excellent thesis, "In Touch with the Wild: Exploring Real-time Feedback for Learning to Play the Violin" [76]. She discusses issues such as demands on focus, variation in student concentration, dealing with shy participants and many contextual issues that, though not core to research, impact results. An important finding is that while valuable, the many contextual variables faced in 'in the wild' studies will yield less definitive results than a similar laboratory study [168].

Johnson also discusses the role of the researcher in cases similar to the one in this study, where the researcher is an active part in the study experience [74, 76]. She discusses ways a researcher directly participating in a study will inherently affect the study beyond traditional research bias. There are issues such as the balance between encouraging participants to engage with the targets of the study while also letting them react freely, the researcher as an authority figure, and the researchers familiarity with the participants. Still, she argues that the opportunity to engage directly with participants in a prototype environment enables gathering a more unexpected, in depth understanding of the target experience.

Though unquestionably important for understanding results, for purposes of brevity and focus, in this thesis, we refer the reader to Johnson's work for in depth discussion of many of the issues and complexities faced during 'in the wild' studies. Though this study uses one-on-one lessons, the fact Johnson's work is based on studies with both children and adults learning violin means much of the discussion remains highly relevant.

9.4.2 Participants

This study was completed by 12 beginners two of whom were adults. We wanted to ensure that students were capable of performing pieces included in the study and have sufficient experience to understand and respond to guidance on pitch. We did not want students using vibrato based on results from the expert study which showed the fingerboard sensor did not always handle vibrato well.

Participants were primarily recruited through two sources, London's Suzuki Hub run by Kate Conway, and students from St. Matthew Academy in Blackheath taught by Sigurd Feiring. One additional adult participant is an acquaintance of the author's and interested in continuing previous violin studies. Students from the Suzuki Hub were recruited from the pool of beginners, including parents Conway deemed appropriate, including parents with children at the hub who have previously studied violin. Students were offered four free 30 minute lessons in return for taking part in the study. Recruitment from the Suzuki Hub also included an introductory meeting where the author presented the augmented violin to students and allowed children and parents to ask questions about the study in advance. Lessons with Suzuki Hub students were conducted at the Suzuki hub with parents present.

The participants at St. Matthew Academy were selected by Sigurd Feiring as he considered them to have an appropriate level of skill and commitment. Normally the participating three students have two 30 minute group lessons each Wed. Due to the time constraints, lessons with students at St. Matthew Academy were targeted for 20 minutes.

All participants, listed in Table 9.1, had completed the Associated Board Royal Schools of Music (ABRSM) Grade 1 exam or first half of Suzuki Book 1, with the most advanced participant just beginning to use vibrato. The study group consisted of 2 adults and 10 children, aged between 31-33 years and 8-11, averaging 32 and 9.3 years respectively. Length

Student ID	Age	Sex	Years Playing	Suzuki	Grade/ Suzuki Bk.	Normal Violin Size	Study Violin Size	Uses Markers
B1	10	M	3	N	2	$\frac{1}{2}$	$\frac{1}{2}$	Y
B2	33	M	1.5	Y	1	$\frac{4}{4}$	$\frac{4}{4}$	Y
B3	9	F	2	Y	1	$\frac{1}{2}$	$\frac{1}{2}$	Y
B4	9	M	3.5	N	3	$\frac{3}{4}$	$\frac{1}{2}$	N
B5	8	M	4.5	Y	4	$\frac{1}{4}$	$\frac{1}{2}$	N
B6	13	M	7	Y	3	$\frac{3}{4}$	$\frac{4}{4}$	N
B7	8	M	3.5	N	1	$\frac{1}{2}$	$\frac{1}{2}$	Y
B8	8	M	4	Y	2	$\frac{1}{2}$	$\frac{1}{2}$	Y
B9	8	F	1	Y	1	$\frac{1}{2}$	$\frac{1}{2}$	Y
B10	8	M	3.5	Y	1	$\frac{1}{2}$	$\frac{1}{2}$	Y
B11	31	F	1	N	1	$\frac{4}{4}$	$\frac{4}{4}$	N
B12	11	F	3	Y	2	$\frac{1}{2}$	$\frac{1}{2}$	Y

Table 9.1: Participants in the beginner study and information about age, experience, whether they were playing their normal violin size, and whether they normally used finger tapes to assist with intonation.

of study was between 1 and 7 years (average 3.7 years) with the grade or Suzuki Book ranging between 1-4 (average 2.2). Half the participants were at a level equivalent to Grade 1. 8 of the 12 participants were female, including one adult. One of the adults is a professional percussionist but all other students have only beginner musician backgrounds. The study group included one severely autistic child who also reported having perfect pitch.

An added challenge for many students in the study is that they were sometimes playing a different size violin or were used tapes or stickers on the fingerboard to provide visible or tangible marking for finger placement when we did not use any in the study. Teachers at both locations agreed that all students should be capable of playing without physical markers though they might find it disorienting and have more difficulty playing in tune.

Lessons were conducted between February 5th and April 25th 2016 with the bulk of lessons completed by March 18th. Lessons were held once per week across consecutive weeks except when interrupted due to school term breaks, events, or illness. All but one student completed at least four lessons with two lessons repeated due to system problems during an earlier lesson. Two students who initially signed up to partake dropped out after one lesson and are not included in results.

Suzuki vs. Traditional Teaching

75% of study participants were Suzuki students compared to 25% that were non-Suzuki. There are substantial differences in how Suzuki lessons are taught compared to most traditional methods. The biggest difference is that Suzuki students spend their first few years learning primarily by ear and also focus on technique from the very beginning. While this meant students generally had good fundamental skills playing the instrument, could sing pieces well demonstrating understanding of pitch, and often had pieces memorized, even the most advanced Suzuki students in the study had extreme difficulty reading music notes and rhythms. A student might know what finger they would use to play a note and even play a complete piece correctly in the wrong key, but could rarely identify what the name of the note was or where it was located on the staff.

Suzuki training affected how lessons were conducted. The author/teacher found that for any piece a Suzuki student was not confident on, it was expected the teacher would lead the piece, requiring the author/teacher to play along. Students followed the teacher for substantial note, fingering, and rhythm guidance and attention was strongly directed at the teacher not a score.

In comparison, non-Suzuki students had a far better grasp of note names and all non-Suzuki students could read music effectively enough to read pieces requested of them. Although non-Suzuki students were used to playing with the teacher, they were less dependent on it and more used to playing on their own. The more experienced non-Suzuki students could also more readily identify the name of the note they were playing based on where their finger was on the fingerboard, or what was written on the staff.

Adults vs. Kids

Only two out of the twelve students were adults. Adults were expected to be more articulate, have higher levels of self-reflection and attention, though lower patience with slow improvement and less likelihood of practicing. Adults more easily discussed the experience of playing with the augmented violin. The two adults also used the augmented bow during lessons as they were the only appropriate participants playing full-sized violins.[4]

9.4.3 Lesson Structure

Lessons were designed to maintain a balance between tasks to keep students interested and learning in lessons while ensuring sufficient repetition of tasks to enable effective comparisons. Each lesson, consistent across feedback styles, consisted of the consecutive four parts in Table 9.2.

Three of the four sections consisted of fixed repertoire in order to be able to compare results across lessons and participants more directly. Scales in participants' lessons alternated between G Major and A Major. The common repertoire was always either Bayly's Long Long Ago, a piece shared between Suzuki Book One and the British ABRSM 2016 grade one exam, or Bach's Minuet III. Both can be found in Appendix A. Bach's Minuet III, also Suzuki Book One, overlaps with the repertoire selection from the expert study. The Bach consists of an A section in G major, used in the expert study and a B section which modulates to D major before returning to G major. Though students played the full piece, collated results for the pitch results of a lesson section used only the music's A section so that lesson section results are always limited to a single key. Which piece was played was decided by mutual agreement each lesson and helped prevent students getting bored from always playing the same piece every lesson. It also gave students a level of choice in the lesson.

Tasks in unstructured time were decided based on what the student needed help with which may or may not pertain directly to the study. Due to the shorter 20 minute lessons at St.

[4] B6, did indeed play a full sized violin but is autistic and it was deemed that the extra cables would provide too much opportunity for distraction and trouble.

Task	Repertoire	Key	Target Length
Scales and Arpeggios	G Major or A Major (2 octave)	G Major or A Major	5 min
Common Repertoire	Long Long Ago (T.H. Bayly) or Minuet III (J.S. Bach, Suzuki Book 1)	A Major or G Major (A section only)	10 min
Unstructured Time	Student Determined	—	10 min
Common Repertoire with Full Headphones (DBL HP)	Long Long Ago (T.H. Bayly) or Minuet III (J.S. Bach, Suzuki Book 1)	A Major or G Major (A section only)	5 min

Table 9.2: Structure of study lessons split into sections with task, repertoire, key, and time spent within a lesson section. In this table, key refers only to musical segments included in overall pitch results for that part of the lesson section.

Matthew Academy, unstructured time was usually skipped unless students made it through fixed study portions quickly.

The amount of time and number of repetitions within each section depended on student capability. Stronger students competent at both the scales and the common repertoire might move through the first two sections of the lesson within a few minutes spending most of the lesson in unstructured time. Students unfamiliar with the chosen common repertoire piece might spend most of the lesson on working on it and skip unstructured time.

All students were assigned a lesson for each feedback style during one their four lessons. The order of lessons, aural, visual, both, or neither, was randomly selected within the constraint that lessons were equally distributed. By randomizing lesson order, we sought to nullify potential effects due to system familiarity.

Soliciting Student Feedback

Participants were asked for their reaction to feedback methods by the author/teacher through two sets of questions, one either side of the last section in the lesson, the common repertoire played wearing both headphone ears. Questions were intended to assess whether students benefitted from and enjoyed the additional feedback experience and also to facilitate conversation. Feedback was also discussed within the lesson as teaching situations arose. Planned questions asked after the completion of the main lesson using a pitch feedback aid were:

List 6

1. What does it feel like to use the feedback tool?
2. How did you use the feedback?
3. Did anything surprise you?
4. Did you find it helpful?

After playing with double headphones testing reduced pitch feedback, students were asked:

List 7

1. Was playing with both headphones fun? Why?
2. What was the experience was like?

The last lesson also included questions on the overall study experience. The final study experience questions were either asked by the author, one of the student's regular teachers, or both. We recognize that having the author acting as teacher and leading discussion with the student into the feedback experience may lead to acquiescence bias, the subtle feeling of pressure for study participants to respond based on what they think the interviewer will want to hear. Acquiescence bias may be exacerbated by the student knowing the system was designed by the author/teacher. Additionally, in an effort to solicit insightful response, the author may be more prone to unintentionally ask leading questions. When logistically possible, or when specially requested due to a student being particularly shy or overly suggestable, the student's associated teacher conducted the interview.

Final interviews with all participants from St. Matthew Academy (B1, B4, B7) were conducted by Sigurd Feiring. Participants B3, B5, B9 conducted final interviews with both the author and Kate Conway. The five overall study questions are listed in List 8.

List 8

1. Of the four lesson types, which was your favorite? Why? (Reminder of lesson types: 1) visual feedback with the colored bar and note name, 2) one headphone with a guide pitch, 3) using both, or 4) no technology, just the teacher!)
2. What was your favorite part of using the augmented violin and why?
3. What did you like the least about using the augmented violin and why?
4. If you were practicing, do you think the visual or headphone feedback would be helpful? When and why?
5. What was it like playing at the end with both headphones? Was it harder playing with both headphones on? Why?

9.4.4 Techical Arrangements

The study was conducted using acoustic violins, full and half-sized, with monophonic input as described in Section 5.7.2. Both the LLAV VST from Section 7.1 and the pitch retuning VST were used Section 7.2 to provide pitch estimation and aural and visual pitch feedback. Aural feedback and visual feedback were provided as follows.

Aural Feedback

For lessons with aural feedback where participants wore headphones covering one ear, headphone volume was set to be comparable to the heard acoustic violin. Also, the pitch shifted audio was played through a speaker at low volume so that the teacher could verify what students were hearing and confirm that scale settings were correct. Besides the multiband compression previously discussed, audio used a low pass filter at 784Hz (G5).

The reason for the low pass filter was a due to a physical build issue. The cabling for the fingerboard sensor on the violin was not shielded and caused interference with the violin clip-on microphone. The interference, audible as a high pitched noise, was not distracting when used with a full sized violin, but on the half-size violin, the smaller distance between augmented electronics and the microphone led to substantial audible noise. Adding the low pass filter helped reduce the noise but at the cost of reduced volume of the guide pitch farther up the violin's E-string.

Visual Feedback

Visual feedback was provided using an enlarged version of the pitch graphic presented in Figure 7.3 of Section 7.2.2. Scores were positioned as near as possible to the screen displaying the graphic so that students could switch visual attention between the two easily. However, due to teaching room layouts and furniture, there was frequently both a height difference of at least one foot (depending on student height), and a displacement of at least two feet to the one side.

Quantitive Data Collection

Quantitive data collected was largely the same as in the expert study. As discussed in Chapter 8, audio was hosted within Reaper where both unchanged audio input and the compressed snapped audio were recorded at 44.1kHz.

Time stamped data from the LLAV VST and the retuning VST was logged throughout a student's lesson. Pitch estimation data changed slightly from the expert study due to the change to a monophonic violin. LLAV VST data logged each logging instance consisted of: clock time, audio sample count, hardware only pitch estimate, combined hardware informed audio pitch estimate, the volume (signal RMS), and any raw bow and fingerboard sensor readings. The data is sufficient for augmented performance playback (Section 7.1.5). LLAV VST logging occurred every 512 samples (11.6ms at 44.1kHz).

The data collected from the retuning VST included everything expected necessary for pitch analysis: clock time received from LLAV VST, audio sample count, received volume (signal RMS), received pitch estimate, target snapping pitch, displayed pitch, and pitch snap strength. Logging for the retuning VST was performed every 128 samples (2.9ms at 44.1kHz).

Qualitative Data Collection

As our sample size is small and the time spent with each feedback method is short, we do not expect statistically significant results, but we do expect useful insight into whether the pitch feedback methods are helpful and why, and also whether the current augmented violin works sufficiently well for use in practice. In order to document responses to the student feedback questions from Lists 6, 7, and 8 and capture unexpected interactions, all lessons were recorded on video along with the audio recorded into Reaper. Additionally, we asked all parents and participating adults the questions in Appendix B for insight into a student's learning style, practice habits, and whether they have experience with practice using variants of the pitch feedback methods in the study.

9.5 Data Analysis

Data collected during the study was split between numeric data on as-played pitch, discussion of the student feedback questions, and observation of events during the lessons.

9.5.1 Pitch Analysis

Analysis of pitch error was carried out similarly to Chapter 8, calculating the mean absolute pitch error, the mean pitch error, and the root mean square pitch error (RMSE) for each lesson segment. For pitch played we used the pitch estimate logged every hop (2.9ms) by the retuning VST and error was the number of cents away from the expected target note. As described in Section 8.2.5, pitch error is measured in cents, or linearized pitch. Within this study, common repertoire sections were scored based on the correct key for the repertoire. If the repertoire key is A major, playing a perfect equal temperament $C\natural$ instead of a $C\sharp$ is an error of 100 cents. 100 cents is also the maximum error as scale notes are either one or two semitones away from each other. All unstructured periods of lessons were scored based on a chromatic scale where every semi-tone was included in the scale so that the maximum error is 50 cents. As much of what we are interested in is the pitch correction process, we calculated continuous frame-by-frame pitch error for the duration of a note.

Removing Invalid Pitch Estimates

Prior to calculating error, invalid pitch estimates from the lesson were removed. Invalid pitch estimates were those where the estimate is clearly wrong or the estimate has occurred while the student is not playing, such as during a rest in the music. The only estimates removed as clearly wrong were those where the estimate was more than 25 cents below the violin G-string (G3) and estimates over G6, well above all included repertoire.

In order to remove pitch estimates when the student is not playing, we removed all estimates for which input signal volume was too low. Since audio input levels were determined by a combination of factors difficult to repeat exactly (for instance, rotary dials on inputs, amplifier battery levels, and exact microphone placement), volume thresholds were calculated by taking one third the mean volume of all hops during scales and the introductory piece.

Though audio segments used for volume thresholding may include momentary silences they still provide a reasonable standard for comparing volumes across differing audio input levels. An approximate level is sufficient as small differences in volume threshold do not significantly change results. For instance, reducing the volume threshold by half, in the loudest lesson overall error increases only 3.24%, while in the softest lesson, with a threshold a tenth of the loudest, error increases only 1.01%.

While the pitch analysis for the study with experienced violinists in Chapter 8 attempted to factor for vibrato, that was not necessary in this study. Only one participant, B5, used any vibrato. B5's vibrato may reduce his apparent average level of intonation. Unlike many participants in the pilot study, B5's vibrato did not appear to impact the effectiveness of the hardware pitch estimate.

In Study Algorithm Variation

There was one significant issue with the augmented violin implementation that arose and was addressed during the study. While we rigorously tested the full-sized augmented violin with multiple users, due to a lack of young beginners to pre-test with, the half-size was only pre-tested by the author, an adult. As lessons with the half-sized violin progressed, it became apparent that it was challenging for many users to sufficiently press down the string to trigger the sensor when playing the first finger.

A similar issue occurs on the full-size violin. Close to the nut, which raises the string so that it does not touch the fingerboard, excessive finger pressure is required to press the string down enough to press the sensor. On a full-size violin the problem is only evident within a half-step of the open string so the pitch estimation algorithm was originally programmed to search within just over a semi-tone either side of an open string when no finger was detected.

The same problem exists on the half-sized violin but as the string is shorter, the pitch range affected is larger, extending beyond the half-step to a whole step. As a result, when students used the one finger to play a whole step above the string (A on the G string, E, on the D string, B on the A string, and F# on the E string), the action was interpreted as an open string and estimated as a the nearest pitch within the algorithm search window for an open string, 70 cents flat of target. As the issue was intermittent it was not immediately obvious whether it was a system or player issue. After a number of lessons using the half-sized violin, it became clear that it was common for many users. The error was easily identifiable when using visual feedback though with aural feedback, it was less conspicuous causing the pitch corrected audio of the specific notes to be out of tune.

While the augmented half-sized violin sometimes struggled to reliably identify low notes on the G string, the failure mode was easy for students to identify as the visuals flash or the audio burbles. However the failure when playing the first finger was subtle and potentially undermined trust so it was fixed part way through the study by expanding the open-string search range to include the whole-tone above. This issue persisted for exactly half of the lessons. and the impact varied by participant and by lesson. Numerical results in the study are all based on the revised algorithm but participant experiences may be affected by which algorithm was used during their lessons.

9.5.2 Video Annotation

Video of all lessons was recorded and annotated, with answers to questions in List 6, List 7, and List 8 transcribed. Annotations were intended to capture interesting discussion or events along with how the lesson was conducted and how a student plays. Annotations were not meant to be exhaustive, but rather illustrative. We annotated lessons for the following set of events:

List 9

1. Lesson Section per Table 9.2.
2. Student playing.
3. Teacher playing along with student.
4. Clear gaze during performance.
5. Explicit focus requested on a non-intonation task such as bowing.
6. Discussion relating to feedback mechanisms.
7. Discussion based on List 6, List 7, or List 8.
8. Other item or event of interest.

We selected each annotation for a different reason. Capturing when the student is playing with the teacher is insightful for two main reasons. Regardless of which feedback style is being tested, if the teacher is playing with the student, the student has some kind of an aural guide. This is not ideal for the test scenario but was often necessary hence, it was useful to keep track of. Second, the teacher playing with the student as an aural guide is much of the inspiration behind the aural feedback case and is also closely related. Tracking it helps us better understand how and when an aural guide is traditionally used.

Gaze is annotated to capture whether or not a student is clearly watching visual pitch feedback tools in an effort to confirm students' self-reported use of, or reaction to visual feedback. Additionally, capturing gaze illustrates attention demands that may conflict with use of visual feedback. Gaze is only annotated during performance when the direction of gaze is clearly distinguishable and sustained. Besides the inherent difficulty in capturing subtle or quick changes in gaze without close target proximity to the camera, the ability to determine gaze is impacted by sight lines, distance to video camera and where items or people in the room are placed.

Periods where students were asked to explicitly focus on non-pitch related tasks are annotated as redirecting a student's focus away from fingering is likely to negatively affect intonation. Lessons included work on posture and bowing where students were told poor intonation was okay. Removing these periods from pitch scoring may be appropriate in cases where intonation was noticeably impacted.

Discussion of feedback tools based on our study questions towards the end of the lesson was annotated. Other comments or questions relating to the feedback tools during the lesson are also annotated to capture how the student is engaging. Lastly, potentially interesting events, such as a student removing headphones, or the author snapping keys incorrectly are annotated as they directly relate to user experience and may impact pitch error results.

9.5.3 Thematic Analysis of Qualitative Feedback

Qualitative results were derived using thematic analysis [16]. Having extracted information on which feedback method was a student's favorite (in the context of their lesson experiences) and which they thought they might prefer during independent practice, all annotations of conversations and events were reviewed for common themes. We conducted two levels of review for themes; the first was themes within or specific to a given feedback method, and the second was themes shared across feedback methods. Comments clearly about a specific single feedback style, such as aural only, or visual only, made during a lesson using combined feedback were considered relevant to the particular feedback method, rather than only in the context of mixed modalities.

Once potential themes were collected, they were reviewed for clarity and significance. Clarity was required not just whether the particular theme was well defined, but whether the meaning of the intention of the student's comments contributing to the theme were sufficiently clear and independently generated. If a question was deemed too leading or pressured, the student's response was thrown out. Significance was given to themes either repeated by a large number of students, or generated by a strong response by a small number of students. Themes are predominantly discussed in Section 9.8 with primary reactions to feedback methods given in Section 9.6.2.

9.6 Results

Of 14 initial participants, 11 completed all 4 lessons with 1 student, B7, only failing to complete the control lesson. As B7 experienced and completed interviews on all three different feedback methods, his results are included where applicable giving results for the 12 participants listed in Table 9.1. 49 lessons were completed with 47 lessons included in overall results. One lesson had to be repeated due to mistakes with study execution, and one lesson was repeated in order to compare two audio input settings. Roughly half of the numerical results of one of participant B6's lesson were excluded from overall study results as B6, who is autistic, would sometimes wander, twirl, or jerk the violin around causing connectors to become loose resulting in a loss or degradation of sensor signal accompanied by a corruption of pitch estimates.

On the whole, 32% of lesson time was spent on scales (18%) and arpeggios (14%), 22% on common repertoire, 38% on unstructured time, and 8% using both headphone ears to test pitch simplification. In a typical lesson, students would play three scales, followed by three arpeggios. As arpeggios were less familiar, these were often repeated additional times. The common repertoire piece was frequently repeated unless a student displayed a high level of proficiency playing it. The double headphone section only featured the common repertoire piece played once unless there was a technical reason to repeat it. 9% of lesson time was spent explicitly focusing on non-pitch related tasks like bowing or posture where focus on intonation was either irrelevant or distracting.

For the context of this chapter, unless stated otherwise we focus on performing correct pitch. Comments on progression, improvement or comparisons in performance refer solely to pitch performance, not rhythmic accuracy, bow technique, posture or other performance tasks.

9.6.1 Quantitative Results

Numerical analysis of pitch accuracy was segmented per lesson section separated according to Table 9.2 between scales and arpeggios, an introductory common repertoire piece, unstructured time, and full headphone performance of a common repertoire piece. Overall results for a lesson exclude the section with full headphones as it used a different feedback method than the rest of the lesson. We present results for each section along with the overall result as error during unstructured time is measured against a chromatic scale with a maximum error of 0.5 compared to other sections scored against a diatonic scale where there is a maximum error is 1.0. Additionally, the tasks represent progression through a lesson allowing change over time to surface in amalgamated results.

The means in Table 9.3 are derived from the means of each participant's mean absolute error per section with repeated scales, arpeggios, or pieces grouped together within each section. The overall error for the lesson was taken as the average across each section.

As part of our expectation is that different participants will respond to different feedback mechanisms differently, we include performance results on an individual basis for each section in Figure 9.1. All participants included in figures completed all four lessons except B7 who did not complete the control lesson with no feedback.

We have also computed statistical significance for the different feedback types per section and overall. All statistical significance testing was done using a one-way ANOVA to test the likelihood of results for mean pitch error using different feedback methods coming from populations with different mean values. Each table includes p-values for whether two feedback forms have statistically different mean performance along with the likelihood based on avail-

		Feedback Method					
		Aural	Aural & Visual	Visual	None	Dbl HP Full Snap	Dbl HP Half Snap
Overall	MAPE	19.62	18.59	18.39	**18.14**	-	-
	STD	(5.25)	(2.50)	(3.47)	(4.27)		
	Low Err	2	3	4	3		
Scales & Arpeggios	MAPE	22.96	22.20	20.37	**19.97**	-	-
	STD	(5.66)	4.38)	(6.26)	(6.07)		
	Low Err	3	3	3	3		
Common Rep.	MAPE	17.75	**16.56**	17.63	16.77	17.21	17.00
	STD	(6.28)	(2.12)	(6.55)	(5.93)	(5.66)	(6.19)
	Low Err	2	4	4	2	-	-
Unstructured Time	MAPE	18.01	**17.03**	17.06	17.64	-	-
	STD	(3.37)	(2.88)	(2.88)	(2.38)		
	Low Err	2	3	4	3		
Preferred Lesson Type		**7**	2	2	1	-	-
Preferred Practice Type		0	**8**	2	2	3	-

Table 9.3: Results of study with beginners using augmented violin with variable feedback. MAPE is the mean absolute pitch error with STD the standard deviation between participants. Low Err is the number of participant for whom the given feedback method resulted in their lowest mean absolute pitch error for that section. Numbers for preferred lesson or practice represents the number of participants selecting the given feedback method as their favorite. Dbl HP is an abbreviation for wearing full headphones blocking acoustic pitch feedback from the violin itself.

able data that the method of feedback makes a statistical difference on performed error. In all three lesson sections, the one-way ANOVA was calculated using the overall mean absolute error for each participant for each lesson section.

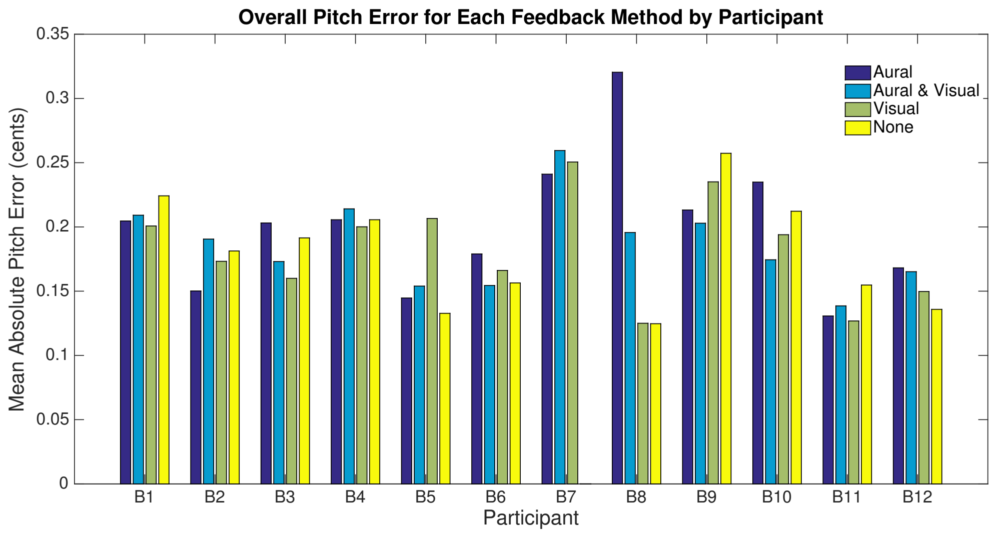


Figure 9.1: Mean absolute pitch error for each feedback method for each participant.

Overall

Lessons with no feedback, as detailed in Table 9.4, had the lowest overall error with a mean absolute pitch error of 18.14 cents while lessons with aural feedback only performed the worst, with error 8.2%, or 1.48 cents higher. Visual and combined aural and visual feedback lessons scored marginally worse than lessons with no feedback, with 1.4% and 2.5% more error respectively. One third of participants had minimal error when using visual feedback. With limited samples and results between participants varying widely, as shown in Figure 9.1, different feedback methods fail to show any statistical significance for affecting pitch accuracy.

Statistical Significance p-values: Overall

	Feedback Type				Overall
	Aural	Aural &Visual	Visual	None	
Aural	-	0.94	0.80	0.64	-
Aural & Visual	0.94	-	0.99	0.94	-
Visual	0.80	0.99	-	0.99	-
None	0.64	0.94	0.99	-	-
Overall	-	-	-	-	0.67

Table 9.4: p-values for differences in mean absolute pitch error between different feedback methods and the overall p-value for whether feedback method suggests any statistical significance whether difference in lesson type is likely to influence pitch error.

Scales and Arpeggios

For scales and arpeggios performed at the start of a lesson, no feedback saw the best average performance with a mean absolute pitch error of 19.97 cents closely followed by visual feedback only. Feedback methods using aural audio guides performed worse, with both having more than 2 cents more mean error, an increase of over 11% than either lesson without aural feedback,. Still, as can be seen in Figure 9.2, variation between players is wide meaning with our limited sample size, results do not demonstrate any clear statistical significance between feedback types (Table 9.5). Despite higher mean pitch error in lessons using any kind of additional aural feedback, 50% of the participants had their best scale and arpeggio performance in lessons using aural or combined aural and visual feedback.

Common Repertoire

During the section where students play one of the two common repertoire pieces differences in error between different modalities were slightly reduced with only 1.16 cents between the

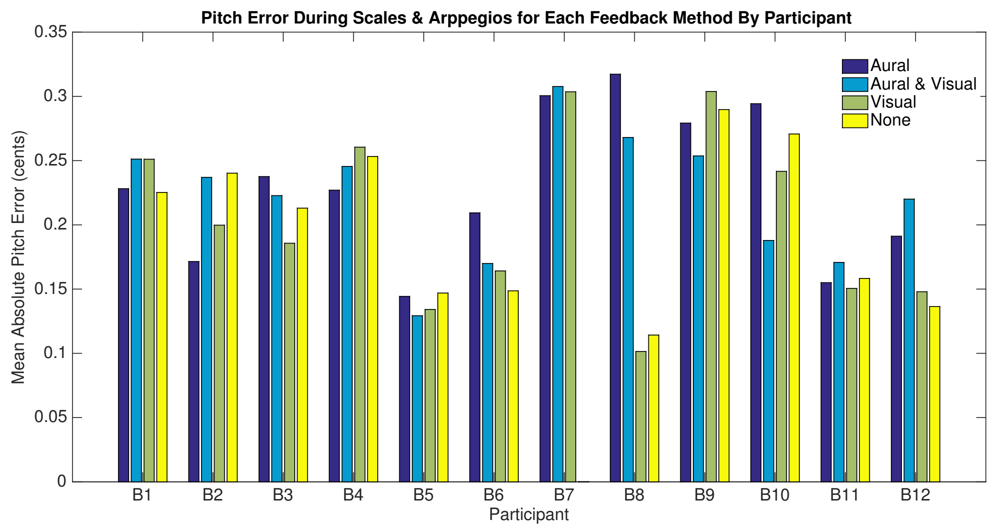


Figure 9.2: Normalized mean absolute pitch error for each feedback method for each participant while playing scales and arpeggios. This lesson section is scored against either a G Major or A Major scale depending on the scale key chosen for the lesson.

Statistical Significance p-values: Scales and Arpeggios

	Feedback Type				
	Aural	Aural &Visual	Visual	None	Overall
Aural	-	0.98	0.63	0.76	-
Aural & Visual	0.98	-	0.85	0.93	-
Visual	0.63	0.85	-	0.99	-
None	0.76	0.93	0.99	-	-
Overall	-	-	-	-	0.62

Table 9.5: p-values for differences in mean absolute pitch error between different feedback methods and the overall p-value for whether feedback method suggests any statistical significance during performance of scales and arpeggios.

highest and lowest pitch error. Lessons with no feedback and both aural and visual feedback resulted in similar error with 16.77 and 16.56 cents mean absolute error respectively. No feedback saw slightly higher error, 0.21 cents higher than with combined feedback, while aural feedback again saw the worst pitch performance with a mean absolute error of 17.75 cents. Interestingly, visual and aural feedback resulted in the most stable results with a standard deviation of only 2.12 cents, half the deviation of any other diatonic performance case. As illustrated in Figure 9.3, one third of participants performed their best common repertoire piece during the lesson using visual feedback only and one third using combined feedback. Again, as Table 9.6 shows, there is no demonstrable statistical difference between feedback styles.

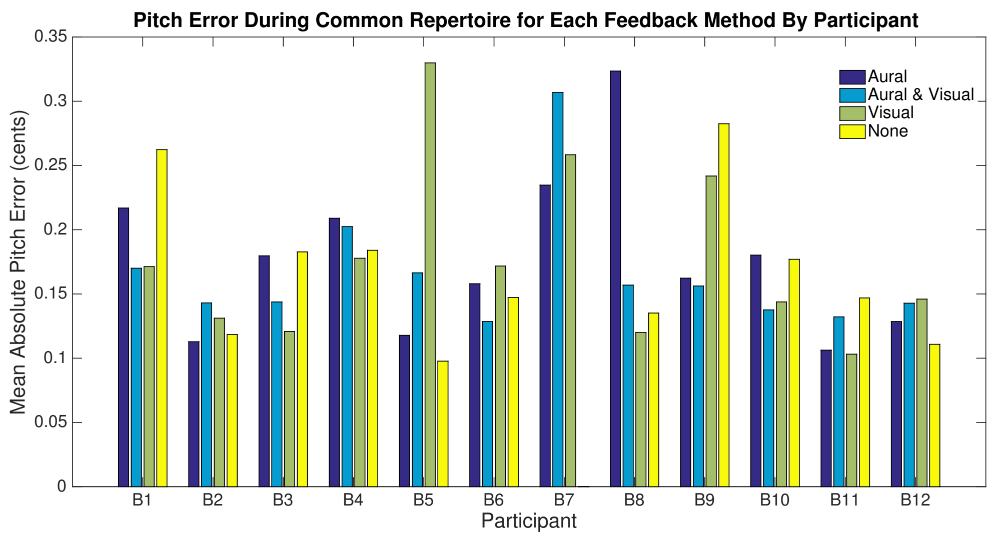


Figure 9.3: Normalized mean absolute pitch error for each feedback method for each participant while playing the fixed common repertoire. This lesson section is scored against an A major or G Major scale depending on repertoire choice.

Unstructured Time

Pitch error analysis during the unstructured lesson time was performed with respect to the nearest chromatic pitch rather than a diatonic scale, reducing the maximum possible error

Statistical Significance p-values: Common Rep.

	Feedback Type				Overall
	Aural	Aural &Visual	Visual	None	
Aural	-	0.84	0.99	0.99	-
Aural & Visual	0.84	-	0.90	0.92	-
Visual	0.99	0.90	-	1.00	-
None	0.99	0.92	1.00	-	-
Overall	-	-	-	-	0.85

Table 9.6: p-values for differences in mean absolute pitch error between different feedback methods and the overall p-value for whether feedback method suggests any statistical significance during performance of the common repertoire pieces.

to 50 cents, but also making it more likely the target note considered in-tune is incorrect. Having both aural and visual feedback yielded the lowest mean absolute pitch error, 17.03 cents, essentially the same as when visual feedback was used (17.06 cents). Aural feedback was again worst with an average absolute error of 18.01 cents, 5.8% or 0.98 cents worse than the best performing lessons. Cases with no feedback scored closest to aural feedback with only a 0.37 cent improvement over aural feedback and a 3.6% or 0.61 cent decrease in performance over visual or combined feedback. Again, as visible in Figure 9.4, lessons using visual feedback saw the best performance for one third of participants. As shown in Table 9.7, we did not find any statistical significance between results.

Reduced Feedback with 'Double Headphones'

Against expectations unlike with the study with experts, as shown in Table 9.3,using headphones with both ears to block the participant hearing the acoustic violin resulted in better as-played pitch accuracy than two cases, aural and visual, where users had normal acoustic feedback. Fully snapped common repertoire pieces averaged 17.21 cents mean absolute error with half snapped cases performing only 21 cents better. Fully snapped double headphones

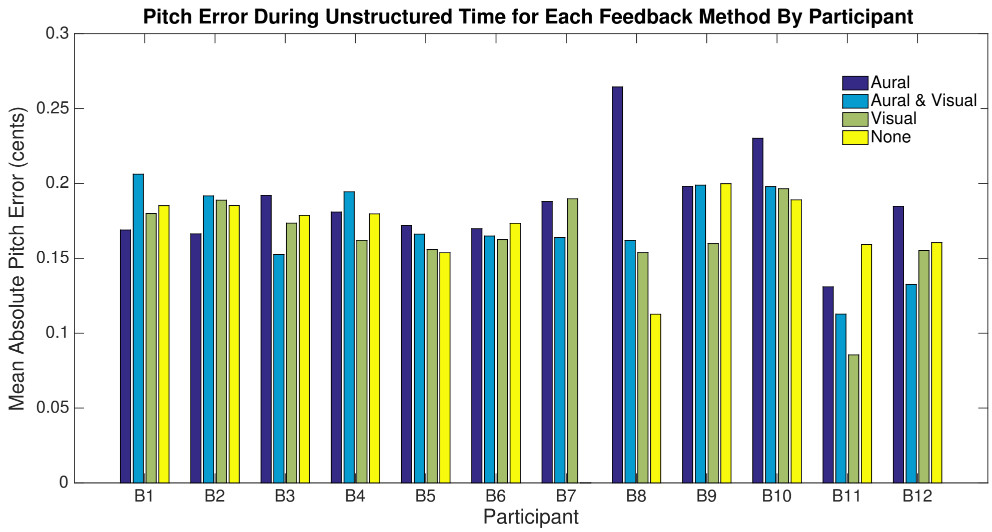


Figure 9.4: Normalized mean absolute pitch error for each feedback method for each participant while studying repertoire based on the student's current progression. This lesson section is scored against a chromatic scale.

Statistical Significance p-values: Unstructured

	Feedback Type				Overall
	Aural	Aural &Visual	Visual	None	
Aural	-	0.83	0.72	0.99	-
Aural & Visual	0.83	-	0.99	0.94	-
Visual	0.72	0.99	-	0.87	-
None	0.99	0.94	0.87	-	-
Overall	-	-	-	-	0.70

Table 9.7: p-values for differences in mean absolute pitch error between different feedback methods and the overall p-value for whether feedback method suggests any statistical significance during unstructured lesson time.

were only 3.9% worse than the best case aural and visual feedback, and better than aural feedback and visual feedback.

9.6.2 Qualitative Responses: Student Comments

While quantitative results are statistically limited, we obtained a wealth of qualitative reaction through annotated casual discussion, questions from List 6, and List 7 asked by the author during lessons and through responses to List 8 collected at the end of the study by either the participant's associated teacher, the author, or both. Overall results presented in Table 9.3 include participants' stated preferences for different feedback types.

Despite the fact that aural feedback saw the worst mean pitch performance during every lesson section, 58% of participants reported aural feedback as their favorite lesson type. Only one student reported preferring no feedback with two opting for visual and combined feedback each. Although not intended as an option, as listed in Table 9.3, three students initially expressed that they found using double headphones the most helpful feedback arrangement.

While the question about favorite lesson type seemed to lead to single modality answers, when changing the question to what a student thought would be most helpful during practice, two-thirds of participants stated a preference for having aural and visual feedback. Three participants reported they would probably only use visual feedback during practice, with one (B1) explicitly saying he did not like the aural feedback as he found it distracting and did not trust it "It helped when I got the 2s a bit off, but when I did the song just now, sometimes it just played notes, random notes, when I played." Two participants reported they would probably not use any pitch feedback method for use while practicing, even though one of them did express that he liked the aural feedback.

Supporting Statements

In all interviews, students were asked to justify and explain their preferences. Here, in Table 9.9 we provide the primary responses to Question 1 from List 8, asking about favorite feedback in a lesson (L), and Question 4 asking whether feedback methods would be helpful during

	Participant											
	B1	B2	B3	B4	B5	B6	B7	B8	B9	B10	B11	B12
Aural		L	L	L	L	L	L			L		
Aural & Visual		P	P	P	P	P		P	L		P	L, P
Visual	P							L	P		L	
None	L						P			P		

Table 9.8: Participant stated preferences for favorite lesson type and what feedback method(s) they speculated they would most like during individual practice. L means participant responded the given feedback method was their favorite lesson style, while P means given feedback method was their preferred option for potential use during practice.

practice (P).[5] If a participant's comments are similar for both questions, or particularly brief, only one answer has been provided. Similarly, if a response was overly long, it has been edited for brevity.

Table 9.9: Headphones refers to aural feedback wearing a single headphone.

Student ID	Ques.	Ans.	Why
B1	L	N	"I liked the person because the person can tell you exactly what you did wrong but computer and headphones they can't tell you."
	P	V	"I only liked the visual because the visual actually helped you see ... where to position your finger... if it's too sharp too flat."

Continued on next page

[5]Questions were not always asked word for word the same as written in List 8.

Table 9.9 – continued from previous page

Student ID	Ques.	Ans.	Why
B2	P	AV	“I would probably be more inclined to use the headphones, just so I don’t have to worry about what’s happening on the screen. All I have to worry about is, is that the right note. .... with the headphones it sort of tells you that’s the wrong note. it’s very indicative that’s the wrong note...probably will need more getting used to... On a new piece I’d need to be more used to the headphones. Let’s say I’m reviewing the pieces, I think, it’s helpful more when I’m reviewing.” A: “And Visuals?” “S: Yeah you can just stare at them actually. Probably more useful than the headphones. The visuals are quite helpful with the pieces I know because I know they are in my head, I’m not reading them, I’m just checking my intonation, I’m checking my fingers.”
B3	L,P	A	“Because you could hear your violin and then you could also hear the tuning [what you were aiming for].”
B4	L	A	“The headphones gave you sort of in tune feedback and mostly when I was out of tune it kind of buzzed, so that helped.”
	P	AV	I think both of them would be equally helpful. It’s because they give you feedback and they give you true feedback, they don’t always give you positive. All the feedback was helpful in its own way so there was nothing that actually didn’t help me. You realize how much you go out of tune and so it helped me to stay in tune.
B5	L	A	“It’s good to have... um... to have the thing that shows you if you’re a bit flat or sharp. Sometimes [the visuals] were [helpful], but sometimes they were a bit distracting. The headphones were good. It’s good to hear the fact that you’re a bit out of tune.”

Continued on next page

Table 9.9 – continued from previous page

Student ID	Ques.	Ans.	Why
	P	AV	“I think I quite like the fact that you can look at it and see, oh, I’m a bit sharp, or a bit flat. Even if it is just slightly and you can’t hear it and it is good for getting it really in tune to perfect a piece. And I also quite like the fact with the headphones I think I’d use them as well because you sort of hear a difference between your tone and the right tone and that sort of helps as well ... and that’s because I prefer to hear it rather than see it.”
B6	L	A	“I like the headphones better because I like to hear what I say.”
B7	L	A	“Headphones because they make the violin sound wet.”
B8	L	V	“Visuals flashy thing because it’s cool. It moves up and down. ”
	P	AV	Just because I like the visuals that’s why I like the visuals. And it is helpful as well because when it’s red you know the options. If you didn’t know which one was which, if you take away the headphones I would see, if you take away the visuals I would hear from the headphones. Both of them is cool.
B9	L	AV	“[I] could see and listen at the same time.”
	P	V	“The visual at all times. I could compare my sound with the [graphic] best, when instead the headphone requires [me] to produce sounds first in order to hear anything.” [Like aural feedback, visual feedback follows the student’s playing. The implication that visuals precede sound is either miscommunication or a misunderstanding of what was happening.]
B10	P	N	Having initially responded none, A: “Why wouldn’t you use it?” S: “Because the flashing thing is just annoying.” A: “What about the headphones? Would you use those at home?” S: “Mmm... Sometimes, but not all the time.”

Continued on next page

Table 9.9 – continued from previous page

Student ID	Ques.	Ans.	Why
B11	L	V	"Think I liked [visuals] the most. It's enough information for me to correct it... like I knew I was out of tune when you were just teaching me without anything, without any aids, but I didn't know how to correct it, whereas this is really clear how to correct it. When I could hear what I was supposed to be playing at the same time, with both the visual and just the headphones, I could hear that I was out of tune and that made it quite stressful, where as [with visuals], there is only one thing I'm listening to, it is just my note which is easier to process and really clear ... whether I am too high, too low.
	P	AV	"Okay, so visually if [the graphic feedback] was more in line with the music... or if it was somehow a program with the score where each note had my intonation on it that would be useful. The hearing thing .. I think.. the reason probably I prefer [the visuals] is that the hearing thing stresses me out because my ear isn't used to it... and the reason it stresses me is that, with my instrument, percussion, I don't usually have to think about [listening], whereas maybe this actually would be good training for me, for my ear, it would probably be good to not have the visual and try and actually hear what I'm aiming for."
B12	L	AV	"Because they're both very helpful on their own, but I think together it's very good because you can hear what you're playing if the headphones give you the wrong thing, you might think that you are playing off, you've got the visuals to see whether you are playing right or you're wrong or not."
	P	AV	"**I think I'd be more enthusiastic about practice than I am now.**"

Continued on next page

Table 9.9 – continued from previous page

Student ID	Ques.	Ans.	Why

Table 9.9: Responses to Question 1 and 4 from List 8, asking about favorite feedback in a lesson (L), and asking whether feedback methods would be helpful during practice (P). Answers are based on feedback styles A- aural, V- visual, AV- aural and visual, and N- none. If the quote includes a two-person discussion, A denotes the author and S is the student.

Aural Feedback

Common comments on aural feedback both in response to questions from List 8 and question variants from List 6 suggest that many users (see B2, B3, B4, B5 in Table 9.9 for explicit support) found, as intended, aural feedback was useful to highlight error and provided a useful guide for matching. Further, when asked during a lesson if the aural feedback was useful, B5, the most advanced participant in the study, responded:

> Especially with the shifting. Because it sort of tells me the right note, like the equivalent to what I would be playing. Except I'm doing it for a shift so if my shift is out of tune it tells me the note. Which helps a lot.

A number of participants commented that the feedback sounded funny or that wearing headphones was odd (B1, B3, B4, B5, B10, B12, see B6 and B8 in Table 9.9). B11 (see Table 9.9) and B5 suggested that aural feedback could be stressful as one might hear they were wrong but did not always know best way to correct it.

Visual Feedback

Participants liked the visual feedback as it also informed users whether they were in tune (see B1, B5, and B11 in Table 9.9). Some users (B2, B8, see B11 in Table 9.9) expressed that an advantage of visual feedback over aural feedback was that it was easier to understand how to correct the intonation and move their finger in the correct direction. As will be discussed in depth in Section 9.8.3, a number of students expressed dislike that the visuals changed so rapidly, calling them distracting (see B5 and B10 in Table 9.9) typically explaining the issue was due to the rapidity with which it could change and flicker.

A constraint on the use of visual feedback confirmed by participants (B1, B2, B4, B5, B11, B12) was that it required visual attention that was not always available. When asked why he was not using the graphics, B1, responded, "Because if you look at it, you can't see what notes you're going to play also at the same time." B2 confirmed this further saying, "Let's say if I'm learning a new piece I probably won't be able to look at [the visuals to see] if I'm on the right note because I'm [looking at the score] and I'm obviously checking my fingers, and I've already got 2 or 3 things to look at."

Combined Feedback

Participants reported that having both aural and visual feedback was useful as they complimented each other well (see B2, B8, B12 in Table 9.9) when aural feedback was confusing, having visual feedback allowed the participant to clarify what was happening, and having aural feedback was useful when visual feedback was not useable.

Control

Two participants, B1, and B7, expressed a preference for no feedback at all in either the lesson or during practice. B1's explanation can be found in Table 9.9 stating a teacher is better at providing accurate feedback. B7 did not provide justification for his response.

Restricted Feedback with 'Double Headphones'

10 of the 12 participants responded positively to the reduced feedback case wearing both ears of the headphone as used in the expert study. By the last lesson of the study, four participants expressed excitement in advance of using double headphones and three participants suggested they would like to use it in practice. Reasons for liking the double headphones include:

> It forces me to listen to what's happening... I think that it's sort of more direct that I need to listen. [B2]
>
> It cuts you off and it helps you focus on your playing. [B4]
>
> You sound amazing. [B5]
>
> I think it would help for practice in a way that, if you just wanted to hear yourself playing a piece, like, if you didn't want to put on the CD and listen to a piece, and you put on the headphones and let yourself play the piece, as long as you played the right notes, it

> would do the right note, meaning that you sort of hear the piece played perfectly... you can't really hear yourself playing. It sort of helps that you'd hear the piece played well. [B5]
>
> B9 says she preferred playing with both headphones on because with only one on she couldn't hear very well. Both headphones helped B9 concentrate best as they helped silencing the surrounding noises too. [B9]
>
> It was easy. It wasn't being out of tune so long as I played the right notes. [B11]
>
> You know on the E string it sounds like the medieval music you hear, like the flute played in films... it sounds nice. It makes you sound better. [B12]
>
> It lets me know, like if you're playing the wrong note and you can hear and then the next time you can correct it and hear if it changes and if it is still wrong. You might not hear like... how the violin normally sounds, but you can hear the notes and how like what it's supposed to sound like. Yeah I like it. [B12]

As suggested in B12's second quote, a number of participants stated they could tell if they were playing the wrong note even though software was limiting the amount of feedback they received (B2, B8, B11, B12). In fact, three students' suggested they found double headphones the most helpful learning aid and gave it as the interaction they felt most likely to use in practice.

Both participants that disliked or had mixed feelings about the double headphones referenced dislike for not being able to hear what they were actually playing. As B1 said "Um... it's harder playing because, when you have just one of the headphones... you could still hear from one ear. Well then if you have two headphones, you can barely hear what you are actually playing so it was a bit annoying."

9.7 Case Studies

As different people's reaction to the different pitch feedback mechanisms varied in preference and clarity and our results were of limited statistical significance, we include two case studies demonstrating clear use of feedback mechanisms and reactions to them. The first is B11, chosen as she communicated the most explicit intentional use of both feedback methods. Sections of B9's lessons including aural feedback are featured as they suggest numerical evidence of use of aural feedback.

9.7.1 B11: Committed Use of Feedback

With most participants, it was unclear how significantly they were using feedback methods. In contrast, B11 is a professional percussionist who requested to join the study in an effort to improve her pitch skills. Unlike other participants, B11s interest was more to learn pitch than learn violin. She is the only participant who was not actively taking lessons with other teachers during or in recent proximity to the study with lessons predominantly focused on learning the study repertoire piece, Long Long Ago. However, as a competent musician, B11 is well disciplined in effective practice and demonstrated quick adaptation to effective use of the augmented violin. B11 was and clear and comfortable reporting to the author/teacher what she felt was working and what was not, along with whether or not she was genuinely using the feedback.

One of the things B11 specifically asked for in the first lesson with aural feedback was that the author/teacher did not play with her stating, "Can you not play...it's different to what's in my ear. ... I can't hear what I'm doing when you're playing," adding "You're really out of tune with what I'm hearing in my headphone." B11 later explained that with the author playing, she had not just her own audio and the aural feedback to listen to, but also the author's acoustic violin and an out of tune version of the author's violin as it bled into the microphone and ended up pitch snapped according to B11s error. Though we discuss the potential broader impact of excessive audio when the author played with a student later, the author largely refrained from playing with B11 in lessons with any kind feedback, playing with B11 only 10.3% of the time in lessons with feedback aids as compared to 39.7% in the control lesson without any additional feedback.

The elimination of an additional source of feedback along with the concentrated focus on pitch and study specific tasks meant B11 was more like a traditional laboratory study style participant than any other participant. The reduced role of the teacher including the focus on her own objectives rather than what the author/teacher might request made B11's study also closer to a normal practice scenario.

Quantitative Results

Table 9.10 show B11's mean absolute pitch error for the different feedback styles. B11s lesson order was aural, aural and visual, no feedback and lastly visual only. Many participants performed notably worse during their first lesson as they got used to the violin and familiar with tasks. Additionally, A major scales and arpeggios, performed during the second and fourth lessons (in B11s case, aural and visual feedback, and visual feedback) tended to be more difficult than G major with scores for A major

	Lesson Primary Feedback Method			
	Aural	Aural & Visual	Visual	None
Overall	13.07	13.86	**12.68**	15.48
Scales & Arpeggios	15.50	17.08	**15.06**	15.83
Common Rep.	10.63	13.22	**10.31**	14.69
Unstructured Time	13.09	11.27	**8.59**	15.91
DBL HP	14.60	11.18	11.92	12.40

Table 9.10: Mean absolute pitch error during lessons by B11.

scales on average two cents worse.

Despite being both the third lesson and featuring G Major scales (generally considered the easiest scale to play on the violin), B11s worst lesson was the lesson with no feedback, with an overall mean absolute pitch error of 15.48 cents. B11's best overall lesson was the last using visual feedback, scoring 12.68, a 22% or 2.8 cent improvement over the no feedback case. Aural feedback, the first lesson follows as the second best, 0.39 cents worse than using visual feedback. The lesson with aural and visual feedback came third but was still a 10%, 1.6 cent improvement over the no feedback lesson.

Lesson Interactions and Commentary

Here we present selected quotes from B11's lessons that pertain to how B11 reacted to the feedback that does not appear elsewhere. Additional quotes can be found in Section 9.6.2 and throughout Section 9.8. In conversation, the author is denoted with an A. During the first lesson, the author and B11 discussed aural feedback:

> A: "Do you find the aural feedback helpful?" B11: "Yes, because I wouldn't know where to go if I didn't have that. If I notice that I'm playing out of tune, it tells me whether I'm flat or sharp. Even that speed we were just doing, it's quite hard. I need longer to get used to this way of playing and I think it'd get easier."

The second lesson used combined feedback with B11 saying:

> B11: "That's much easier without looking." [Refering to playing Long Long Ago not looking at visuals, just using headphones]
>
> B11: "I guess I wasn't really using visual. In the beginning I thought I was meant to use the visual. I was looking at [the visual] a lot, but when I realised I didn't have to, I stopped using it so much. When I had both headphones on, it was helpful to know that I was hitting those $C\sharp$s especially. And it was nice to see that I was in tune in some bits apart from when you moved my hand for posture reasons, then i was like ...argh..."
>
> A: "How is the aural feedback?" B11: "Good." A: How is it helping you?" B11: "It's really helping, it sounds horrible when it's out of tune, but when I get it's like, ah yeah!" A: "If it's out of tune, do you know which direction you have to go?" B11:"I guess this was when I was trying to take [the visuals] all in. If I do one without visual I feel like it's good. It's easy to quickly correct to follow the pitch." A: "You mean it sounds horrible and you can hear the right note and correct to that?" B11: "No, I mean it sounded horrible when I was using the visual. Too horrible and too confusing to know how to correct and make it sound good. Which is why I'm just using the ear thing. It's easy to correct and it doesn't sound horrible."
>
> A: "Earlier on when I forgot to set the pitches, was there a difference?" B11: "Yes, there was something that happened there. I noticed it because I did the wrong semitone, and what was in my ear was aiming for the wrong thing. Even though I knew it was wrong, the thing in my ear wasn't helping so it was more difficult."

During the lesson with no feedback B11 commented:

> B11: At the beginning of the scale I didn't know where to go, but if I go slow enough I do. There was a bit where I didn't know if I was flat or sharp and then I missed [the aural feedback] ... I admit I missed then having the guide.

The last lesson with visual only feedback included discussion of when it was useful:

> B11: "I was looking at the visual and I knew what I was aiming for." A: "You found the visuals helpful?" B11: "Yeah I did." A: "How often were you looking at them?" B11: "The second time [playing the piece] I was all the time." A: "When were you not looking at them?" B11: "When I was looking at my fingers or the score."

Discussion

Throughout the above and in quotes by B11 elsewhere (Table 9.9), B11 repeatedly referenced the usefulness of both enhanced feedback modalities at highlighting error, sometimes unpleasantly, and providing information for how to correct. B11's stated use of feedback and response that both aural and visual feedback can both be helpful is backed up by her numerical pitch accuracy. Though it would be inappropriate to suggest results are statistically relevant considering the differences in tasks, repertoire/feedback familiarity and concentration on the day, all three feedback methods substantially outscored no feedback, 1.62 cents worse than the closest enhanced feedback method. This is despite the fact that no feedback was the third lesson by which time B11 was better practiced. In that lesson, B11 even stated she missed having the aural guide to help her know how to correct.

As we've stated, though we tested in a lesson context for practical reasons, feedback methods are designed for practice purposes and B11's comments support it's applicability to practice. B11 said:

> "When you weren't giving me a lesson, i.e. stuff about bowing and posture, it really clarified what I was aiming for and what I needed to do, i.e. move my finger up or down the string. When I didn't know what to do, it was frustrating and overwhelming.

In fact, B11 suggested that the aural guide provided through aural feedback might be more helpful than the original it was modelled on:

"I preferred the feedback to your violin: audio and visual. It was much easier to hear and see. Because of the compression, I could tell which one I was aiming for really clearly. When I had got into it, I was getting to grips with the feedback and was getting much better at correcting my pitch."

Interestingly, B11 also shows growing comfort with different methods and lesson material. In her first lesson, B11 remarked she needed to get used to playing with aural feedback. B11 reported preferring aural feedback during her second lesson with both aural and visual feedback. However, in her final interview, after the lesson with visual feedback only, she stated she liked visual feedback best (see List 9.9). By the final lesson, B11 was much more familiar with the common repertoire piece she played, Long Long Ago, confirming that she had memorized the piece which allowed her to use visual feedback more effectively. Increased familiarity with both feedback method and repertoire may explain B11s shift in preference.

9.7.2 B9: Referencing of Aural Feedback

Without relying on participant self-reporting, use of aural feedback can be difficult to assess since listening is not externally visible. When everything is working correctly, it can be difficult to infer whether an improvement in pitch accuracy is due to the enhanced feedback tools or the student's self-assessment based on their internal memory of the piece, teacher instruction, or the student listening to the teacher's aural guide if playing together. However the study did include some episodes where, due to an incorrect key setting for aural feedback, there was evidence that a student followed the enhanced feedback rather than what was correct demonstrating use of aural feedback. Here we discuss one of the more clear episodes.

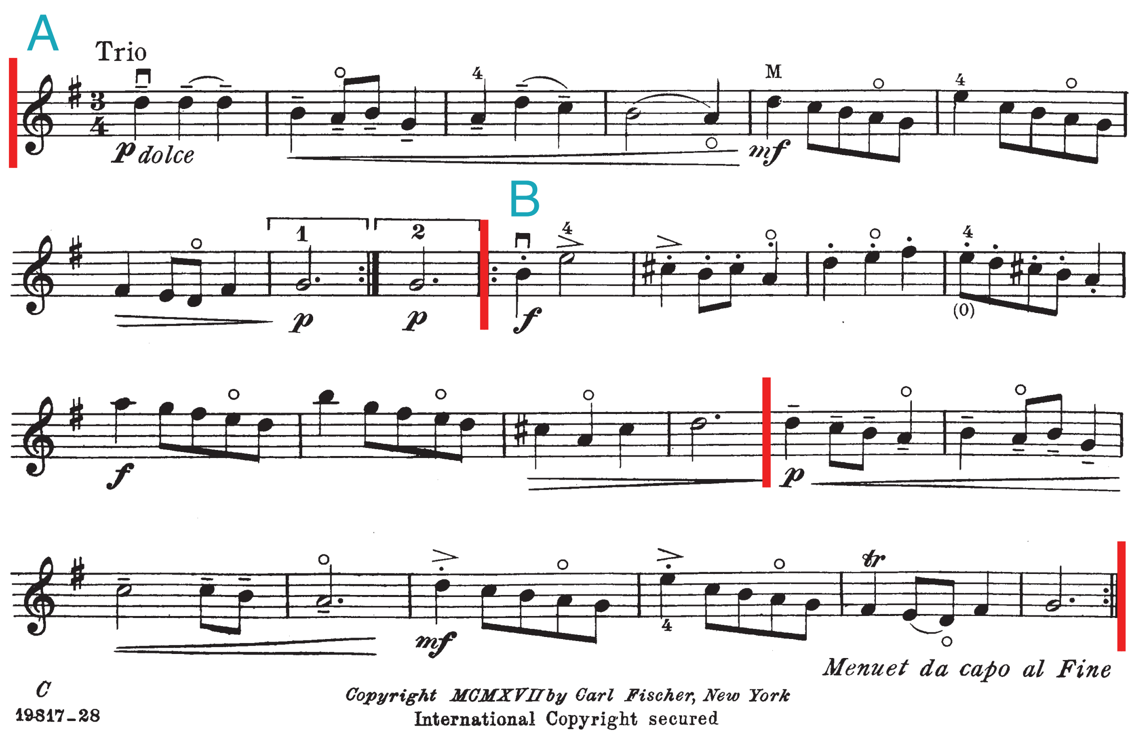


Figure 9.5: Score to Bach's Minuet 1 (arr. Seely-Brown). The piece begins with an A section in G Major. The second section, B, is composed of a section in D Major followed by one in G Major. A red bar marks the change in key along with starts and ends of sections.

During B9's last lesson, which was using aural and visual feedback, having worked on Long Long Ago, in A Major, we switched to Suzuki Book One's Minuet I by Bach, shown in Figure 9.5, which is in

G Major (A section), with a B section that modulates to D Major before returning to G Major. For Long Long Ago, the author had set aural feedback to snap notes to A Major and failed to change the key for aural feedback for the new piece. Due to the wrong key setting, the aural guide incorrectly continued snapping all $C\natural$s to $C\sharp$ and $G\natural$s to $G\sharp$.

B9 proceeded to play Minuet I on her own with a mixture of $C\natural$s, $C\sharp$s and also $G\sharp$s. As she continued playing, pitch error went up significantly so that after playing the the G Major A section with an average error of 27 cents, the next two G Major segments in the piece's B section had a mean absolute pitch error of 42 and 37 cents each. The increased pitch error was largely due to incorrect $C\natural$s and $G\natural$s. Figure 9.6 illustrates the second of these two sections. The expected $C\natural$s (3 steps above A) are closer to $C\sharp$s (4 steps above A) and the intended $G\natural$ (2 steps below A) is performed much closer to a $G\sharp$ (1 step below A). If scored against A Major, to which the aural guide was set and which contains the two additional sharps, both segments' pitch error improve to, respectively, 29 cents and 30 cents mean absolute error.

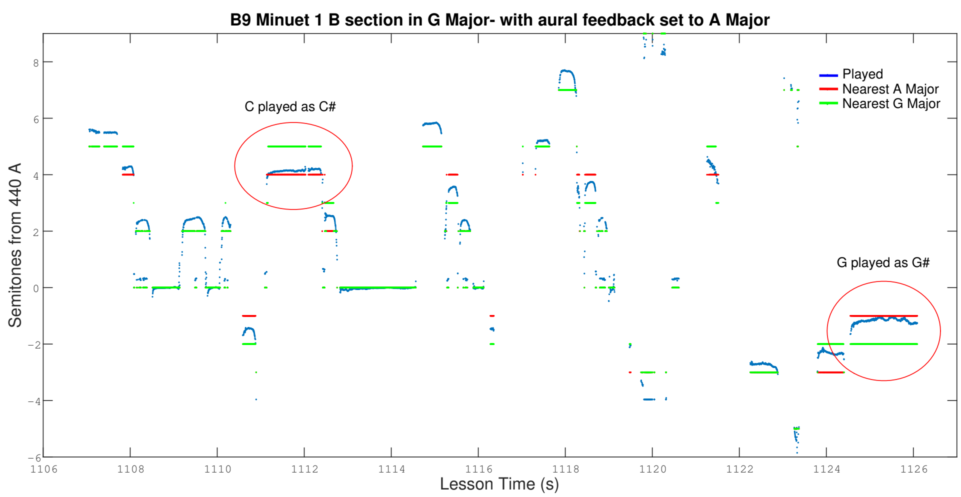


Figure 9.6: Section of B9 playing Bach's Minuet 1 from Suzuki Book One. Aural feedback is incorrectly set to A Major and B9 can be seen using $C\sharp$ and $G\sharp$ used in A Major, but not in G Major. Red lines represent the nearest note in A Major, and green lines the nearest note in G Major. In the case of the circled $C\sharp$, the note should be $C\natural$ but as the played note is sharper than $C\sharp$, the nearest note in G Major is D.

The repeated incorrect $C\sharp$s triggered the author to notice the missing $C\natural$s in the snapped scale but not the incorrect $G\sharp$. After restarting the piece, this time with the author playing along with the student,

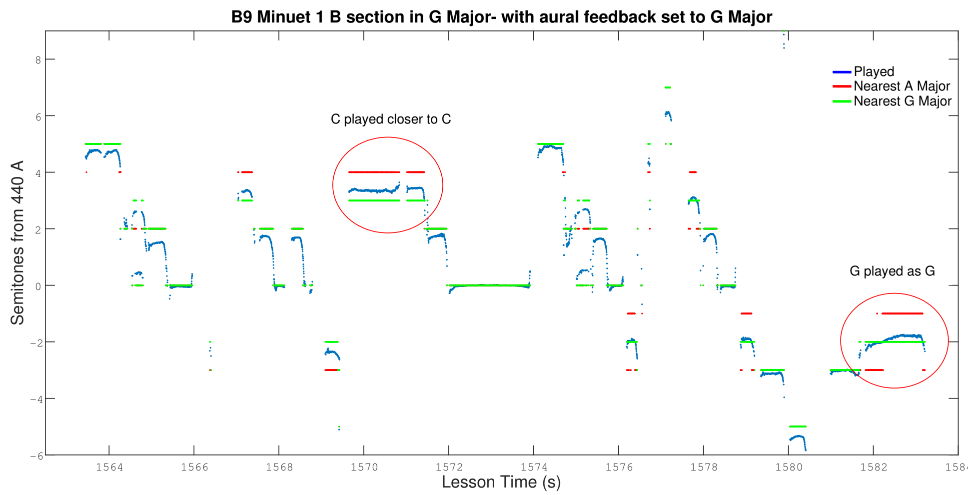
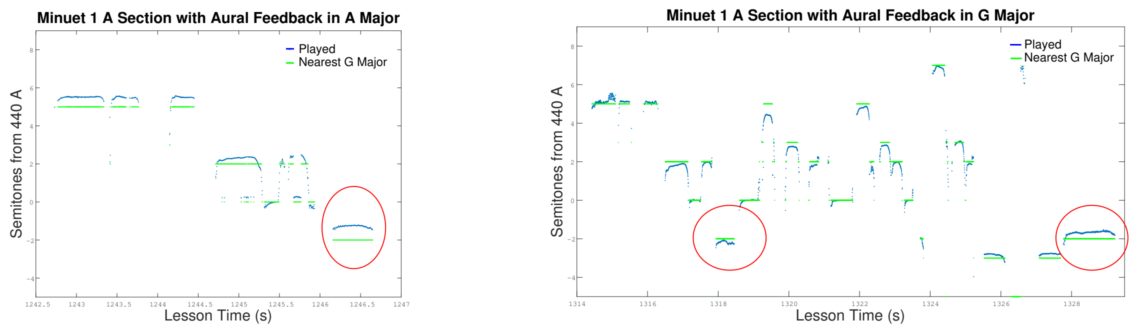


Figure 9.7: Section of B9 playing Bach's Minuet 1 from Suzuki Book One. Aural feedback is correctly set to G Major and B9 can be seen using $C\natural$ and $G\natural$ instead of the $C\sharp$ and $G\sharp$ when snapped to A Major. Red lines represent the nearest note in A Major, with green lines the nearest note in G Major.

though B9 started fixing the $C\natural$s, she continued to play incorrect $G\sharp$s even after the incorrect pitch had been pointed out repeatedly. This being out of character for the student, the author realized the aural feedback was still snapping the student's audio to $G\sharp$.



(a) (b)

Figure 9.8: Section of B9 playing Bach's Minuet 1 from Suzuki Book One with highlighted $G\natural$s just before (a) and just after switching scales (b) from A Major with a $G\sharp$ to G Major.

Upon fixing the aural feedback to the fully correct scale, pitch accuracy improved dramatically. For instance, as depicted in Figure 9.8, in the section of Minuet played just prior to fixing scale settings (Figure 9.8a), the $G\natural$ was roughly 75 cents sharp, 25 cents flat of $G\sharp$. Immediately after fixing the snap, the $G\natural$ improved to 45 cents sharp, 30 cents closer to correct (not shown in Figure 9.8), before further improving to 11 cents flat and 30 cents sharp (Figure 9.8b). The fixed $C\natural$s and $G\natural$s resulted in an improved pitch accuracy of 25 cents immediately after the scale settings were corrected in the piece's A section, and 19 cents the following repeat, an 8 cent improvement over the start with the incorrect guide. Additionally, the pitch accuracy the next performance of the piece's B section, depicted with incorrect guide in Figure 9.6 and with correct guide in Figure 9.7, improved from over 37 cents to 22 cents mean absolute error.

During the previous lesson using just aural feedback, when asked if she was listening to the aural feedback, B9 responded she was "Listening to it." Asked further if she was using it, she confirmed that she was but was unenthusiastic. During the lesson in this case study, she did not express any verbal indication whether she was using either aural or visual feedback methods nor did she comment on any confusing pitch behavior by the augmented violin.

Discussion

Here, through the accident of incorrectly set scales, we can see strong evidence of B9's use of aural feedback as her $C\natural$s and $G\natural$s appear to follow the guide, not her memory of the piece, nor teacher instruction. Participant B9 was quite shy and did not like answering feedback questions. In fact, during the initial final interview, she did not answer more than to say she liked the double headphone experience. It was hard to evaluate how much she used aural feedback and whether she found it helpful, though here we can. It is possible the author/teacher asking her to correct incorrect notes contributed to the improvement, but verbal instructions for correcting pitch were limited and those given were not achieving results. Once the aural feedback was set correctly though, B9 improved rapidly. It is unlikely the improvement in pitch was the result of visual feedback as not only is visual feedback displayed chromatically and not snapped to any scale, but the student also spent the vast majority of playing looking at the fingerboard.

While the error in the aural feedback setting would have increased her pitch error when following, if she was successfully following a correctly set guide at other times, we would expect her pitch accuracy to be better in lessons with aural feedback. Though not necessarily statistically valid, B9's broader results in Figures 9.1, 9.2, 9.3, and 9.4 support that B9 benefited from aural feedback with both aural feedback and combined visual and aural feedback producing the lowest pitch error in all cases except unstructured work.

Outside this case study, further evidence of B9's reliance on listening for pitch accuracy, whether supplied by the augmented violin or acoustic, is that in her third lesson, she performed Long Long Ago for the last lesson section using the double headphones so she couldn't hear herself. Though she successfully performed Long Long Ago as written in A major multiple times earlier in the lesson with only 18 cents error, she unintentionally performed the piece in e minor, using $C\natural s$ with only occasional $C\sharp$s, leading to a mean absolute pitch error of 33 cents nearly doubling the error from when she could hear herself.

9.8 Discussion

Quantitatively it is difficult to make any definitive conclusions whether aural and/or visual feedback are effective for improving a student's pitch accuracy, however user feedback and our case studies suggest that aural, visual, and combined aural and visual feedback may all be helpful learning intonation. Numerical results of pitch analysis suggest that rather than helping, additional aural, or visual feedback hinder correct pitch accomplishment, but that combined aural and visual feedback is the best option when using feedback. As previously stated in Section 9.4.1, ‘in the wild’ studies often lead to more mixed results than laboratory studies, and indeed, during our study we saw pitch accuracy vary due to student practice outside the lesson, concentration on the day, missing finger tapes, an unfamiliar violin size, and more. Additionally, given our small sample size, we would not expect nor were our results anywhere near statistically significant results making qualitative assessment potentially more insightful.

After a note about qualitative bias, we discuss user reaction to aural feedback, visual feedback, and using both together. We also discuss reaction to double headphones, playing wearing both ears of the headphone, before moving onto some of the more universal issues we found such as the effects of trust, whether participants thought feedback was useful or just fun, demonstrated practicality of the augmented violin, highlighting error more aggressively, and the possibility of using compression to simplify bow interactions in future complexity management experiments.

9.8.1 Qualitative Bias

Of course, qualitative responses can be susceptible to bias and unreliable statements. For instance, answers may favor more recent lesson types or be influenced by interviewer comments. Evidence of this occurring within this study is supported by numerous contradictory statements, typically between different lessons. For instance B2, B3, and B11 (discussed in the case study of Section 9.7.1) both stated at different times they liked either aural or visual feedback better than the other only to reverse their preference in another lesson.

Lastly, we assume that students are being truthful. Despite efforts to encourage honesty with statements like “Feel free to be honest. It’s actually more helpful if you are honest even if you don’t like it” we expect a level of acquiescence bias as students generally sought approval from the author as teacher. However in some cases students were a bit more difficult and contrary. For instance, after slowly correcting a number of notes while playing, the author asked B8:

> A: “Are you listening to [the aural feedback] at all?” ...
>
> B8: (confident) “Yeah.”
>
> A: “Is it helpful? Sometimes?”
>
> B8: “No.” ...
>
> A: “So when you are tuning, for instance at the end. Your two was out of tune but then you retuned it. You tuned it right... That was a good thing, but I’m just asking because I’m curious.”

B8: "How did I tune it?"

A: "Yeah. Because you heard it was out of tune and then you fixed it. How did you decide that?"

B8: "Well, maybe the headphones."

9.8.2 Aural Feedback

While aural feedback repeatedly resulted in the worst pitch accuracy for many users, it was by far the most popular lesson type (7 out of 12). Let us look at why students liked aural feedback and also why it might not always have performed as well.

As discussed in Section 9.6.2, numerous participants pointed out that aural feedback highlighted error (B2, B4, B10, B11, B12) and provided a valuable illustration of correct (B2, B3, B5, B11, B12). Students also expressed that they appreciated that it was essentially a passive form of feedback (B2, B4, B5, B11); using it did not require a constant shift of attention. B5 remarked, "you sort of don't really know that it's actually playing, because well when you're in tune it sort of melds into your playing." However when acoustic and electronic pitches clash, it stands out.

Despite largely positive response to aural feedback, a major issue likely weakening the limited numerical results in this study, is that it takes some time to become accustomed to.

Familiarity

A significant issue highlighted by B2 is the need to get used to aural feedback, "I think it's because I'm trying to listen in a different way... I'm [used] to trying to listen to the [acoustic] rather than what is coming through [the headphone ear]". B2 also repeatedly expressed the desire to play more with the aural feedback to "Get used to it". Learning to balance focus between the acoustic violin and the headphone feedback was a distinct change from unaugmented practices.

In fact, strong evidence that aural feedback does indeed take some 'getting used to' comes from further separating lessons between scales and arpeggios. Table 9.11 presents these expanded results encompassing complete lessons. In both methods using aural feedback, scale performance is more than 2.5 cents worse than other methods. This gap is the largest difference between group means seen in this entire study. Further, when moving onto arpeggios, aural inclusive methods both improve by at least one cent while non aural inclusive modalities improve by half that if at all. For the common repertoire section, aural inclusive modalities have improved yet again and are no longer clearly worse than non

	Lesson Primary Feedback Method			
	Aural	Aural & Visual	Visual	None
Scales	23.24	22.81	20.12	**20.05**
Arpeggios	22.01	20.74	20.03	**19.54**
Common Rep.	17.75	**16.56**	17.63	16.77
Unstructured Time	18.01	**17.03**	17.06	17.64
DBL HP	20.23	16.12	18.04	16.89

Table 9.11: Reduced version of Table 9.3 with scales and arpeggios separate. Improvement in pitch error throughout the lesson when using any aural feedback as compared to much reduced improvement in cases without headphones suggests it may take time for a user to become accustomed to aural feedback. DBL HP is a form of reduced feedback different to the main lesson feedback style, but are included for comparison of lesson progression.

aural modalities. Improvement between scales and common repertoire for aural feedback alone is 5.49 cents with combined aural and visual feedback improving a notable 6.25 cents, a hearable difference. In comparison, visual feedback only improves 2.49 cents, less than half either aural inclusive feedback, with no feedback improving 3.28 cents.

Some of this difference may be that, while visual feedback uses a sensory modality which is not strictly required for performance, aural feedback alters an already in use and key mode of feedback. Participants must become accustomed to hearing their violin in only one ear while hearing a compressed pitch corrected version which does not match their playing in the other. While participants are used to playing with teachers and hearing two people playing, participants reported aural feedback was quite different (B8, B11). Unlike aural feedback, the teacher is not the dominant sound in one ear.

Additionally, while audio feedback audio was generally of good quality and free from artifact during playing, it still sounded different. Not only were there times when it burbled on low strings or as a result of poor finger pressure, poor string crossings, or terrible tone but due to the heavy compression, ambient sounds from the room were amplified as well with students commenting “Everything sounds weird now (B8)” or “Everybody sounds like a robot (B3)”. All of these could potentially cause distraction when heard in the headphones. Participants had to get used to ignoring or accepting them.

Attention and Volume

Demanding less constant attention than visual feedback does not mean that aural feedback does not need attention. A common question during lessons with aural feedback was, "Does that note match what you hear in the headphone?" Adults B2 and B11 both admitted at times "I don't think I was listening".

Part of making aural feedback effective and drawing attention is ensuring correct volume. Too soft and the participant can not use it effectively, yet too loud and it becomes painful and distracting. For instance, B11 pointed out in a lesson, "I can't really tell if I'm playing in tune. Can I have the volume up in my headphone?" When pointed out he was not correcting notes higher on the E string, B2 responded that he could not hear those notes well. The loss of volume in higher frequency aural feedback was due to filtering to remove electronic buzz, but this demonstrated the very real practical implications of volume. After removing the low pass filter, B2 corrected the higher notes effectively.

B11 also noted that the low pass filter made aural feedback less effective at higher pitches stating, "Lower down [in pitch] is much easier to hear the difference between what I'm playing and what the headphone is doing ... when I can hear it, I use it to tune."

Additionally, if volume was too high and uncomfortable, the discomfort would out-weigh the positives of aural feedback. Intentionally testing volumes with B2, "I liked that [volume], the interesting thing is I want it to be louder when I play the wrong notes. When it's this loud it was good. Probably if it was louder it would be too much."

We discuss further ways to increase attention in Section 9.8.10, but for now, we point out that ensuring sufficient volume had a significant impact on aural feedback's effectiveness and even then, having aural feedback does not mean it was used.

Sensory Overload

One potential issue with aural feedback in the lesson context especially when playing with the teacher was auditory overload. Auditory overload was pointed out by B11. Commentary on her experience can be found in Section 9.7.1, but she pointed out that when the author/teacher played with her she had too many conflicting versions of audio to effectively respond to. When asked what was her least favorite part of using the augmented violin, B11 replied, “I remember when I had the headphone, and me, and you were playing and that was way too much, and you were triggering wrong notes and stuff.”

Alerted to the issue, the author/teacher responded by playing less in lessons, however based on video annotation, the author/teacher still played with students 47% of the time. In cases where the student expected the teacher to play with them, especially in lessons with Suzuki students, the author tried to play quietly so as not to interfere with or overshadow the aural feedback, however it may have still been problematic.

Rejecting Feedback

Because aural feedback was continuous and unavoidable as long as the student was wearing headphones, it was easy to capture rejection of aural feedback. Participants were told they were free to remove headphones if they wanted, but in the 24 lessons using aural feedback, headphones were removed in five lessons, twice at the author’s behest, and three times at the student’s request. In all cases where headphones were removed, students completed all but the unstructured time wearing headphones.

B3 requested to take headphones off roughly halfway through both of her lessons using aural feedback. She gave a combination of reasons; they were bulky, the cable was annoying, and they were distracting. Despite this, in her final interview, she said she liked aural feedback and found it helpful. Possible insight into why B3, who rarely talked freely about her experiences, might have liked aural feedback but wanted to take off the headphones comes from B4. B4, who liked the aural feedback, would frequently remove his headphones when not playing stating that the background noise, particularly the whine from the electronic interference with the microphone was unpleasant and hard on the ears. In future efforts need to be made to reduce background noise in aural feedback.

Rejection of aural feedback due to the background noise was definitely not the issue for B8. In the lesson using aural and visual feedback, B8’s last lesson, he asked to remove the headphones:

> Can I take the headphones off? It sounds weird... It makes me feel I’m playing the wrong note but I don’t think I’m playing the wrong note, because it always sounds weird... it kinda makes me make a wrong note but I don’t know if it’s wrong or not. (B8)

B8 is also the one participant whose pitch accuracy was distinctly worse in lessons using aural feedback (see Figure 9.1). Part of the issue in the lesson where he took the headphones off turned out to be that the volume as too loud, he still enjoyed wearing both headphones at the end once we had turned the volume down slightly, however incorrect volume is unlikely to differences in performance. More likely, he is an individual for whom multiple audio sources was overwhelming and confusing.

The one of the two times the author requested a student to remove headphones was due to discipline issues, and the other for technical reasons. B6, who enjoyed using the headphones, was using them as a distraction and being unruly so we took them away. The author also told B4 to remove the headphones in one of his lessons when it became apparent that there was a problem with the host computer that was causing excessive glitching that was getting worse.

9.8.3 Visual Feedback

Students generally commented positively about visual feedback with 10 out of the 12 participants saying they liked it and found it helpful. As presented in Section 9.6.2, participants commonly stated that visual feedback was easier to understand than aural feedback, that they had a personal preference for visual feedback (B2, B11), or they simply liked how it looked (B8). The graphic design was successful in that it was praised as easy to interpret the colored block: below the in-tune line means raise the note, and above the line means lower it. For participants with a poorly developed intonation process loop, deciding which way to fix the note through aural analysis was more difficult.

Even considering visual feedback's short term usefulness, it suffers from two major flaws, gaze and speed.

Gaze

In order to use visual feedback, the student must be able to look at it. Section 9.6.2 includes a number of participants remarking on the challenge of including visual feedback while wanting to look at the violin or the score. To evaluate the impact of that, we looked at gaze.

Gaze annotated during lessons, a sampling of user visual attention, suggested during lessons using visual feedback, the graphics did not regularly receive attention. Though indicative of use, our simplified measure of gaze is not perfect, for instance, B4 remarked he watched the feedback out of the corner of his eye while focusing on the score. Of annotated periods in lessons where visual feedback was available, participants were clearly looking in the direction of visuals only 16% of the time. In comparison, 59% of the time participants appeared to be looking at their violin, and 19% of the time they were looking at a score. The remainder of gaze was directed at the teacher, the bow, or around the room.

Participants most commonly looked at visual feedback during scales and arpeggios. Scales and arpeggios, which were taught aurally, did not require use of a score making it easier for participants to direct visual attention at the graphics. They were also closer to the beginning of the lesson, where the author introduces them making the visual feedback fresh in the mind.

As a result of issues with visual attention, the consensus amongst adult participants, B2 and B11, was that visual feedback was only an effective modality of feedback once the notes of a piece had been learned. When starting a new piece, wearing either one or two headphones was far more useful.

Sensory Overload

As noted in Section 9.6.2, visual feedback also suffered from sensory overload, mostly due to feedback changing too rapidly for users to process and flicker due to estimation error between notes. Despite the fact that we were only updating the visuals every 30ms, while sometimes "Calm" (B11), participants still described stated that at times the rate of change as confusing. B10 called the visual feedback graphic the "Flashing thing" saying it was distracting while B11 stated:

> [The visual] is useful. It's interesting. Obviously the colour and the direction of the bar does help, but in a playing situation, it's way too much and it's flickery and it makes me just really confused, where as if i just focus on what i'm hearing it's much easier to play it in tune... it was all red bars everywhere and it was going too fast to correct it whereas if I'd of been able to hear it I could have corrected it.

A challenge designing a useful visual feedback tool is that music proceeds rapidly not always leaving enough time to correct. B2 stated about pitch correction in general, "It's a question of how much time I had to shift my fingers. I don't think I had much time to shift them because I was trying to play the same speed as I was playing earlier."

While for very fast notes, even aural feedback will be unusable, aural feedback has the advantage that we can process auditory stimuli faster. Aural feedback was just over 11ms behind participant performance and it takes only 40μs for the brain to convert sound to neurological signal [65] but audio glitching was not normally overly problematic. In comparison, visual feedback was updated every 30ms but it takes 50ms for humans to trandsuct visual stimuli in normal light [65]. Aside from participants complaint of flicker, even for beginners, note durations are regularly below 200ms leaving very limited time for a user to correct. We have stated already that visual feedback from slower commercial tuners that flicker less are considered too slow to be useful in practice, but simply speeding up the tuner response does not necessarily overcome the problem that human visual processes are slower than aural.

Familiarity

Participants found the basic bar graphic described and pictured in Section 7.2.2 easy to use, but note names were of limited use. Only 3 study participants, B4, B6, and B11, showed competence at naming notes on a score, though B11 was not familiar with where the note was located on the violin. All other students referred to notes by their first position fingering for instance, D4 would be referred to as the third finger on the A string. Even more experienced players displayed poor knowledge of where a named note was on the violin, as well as where the named note was on the staff. This meant that only two participants were able to use the portion of visual feedback naming the note.

The problem with note unfamiliarity is that students can not interpret information from the visual feedback regarding whether they near the correct note or not. Say a participant attempting a $C\natural$ natural played 60 cents sharp, as 60 cents is closer to $C\sharp$ than $C\natural$, the graphic would suggest to the student they were 40 cents flat of a $C\sharp$ and needed to raise the pitch rather than lower it. Visual feedback did not follow a diatonic scale as it was expected that students would understand that a 'high two' on the A string was a $C\sharp$ and a 'low two' is a $C\natural$. This assumption turned out to be mistaken.

However the fact that students have not learned note names provides a learning opportunity when students repeatedly see the note name associated with the note played. B2 and B8 both recognized the usefulness of visual feedback for learning to associate notes played on the violin with note names. B8 stated "Yeah. I would [use the visual feedback] so i know F# and A#. I will know the notes; how to say them in the letters."

In future it would make sense to add the option to display the first position fingering. Adding first position fingering would make visual feedback more understandable to all students and if used in conjunction with note names, could potentially help speed up the process of learning the names of notes.

9.8.4 Aural & Visual Feedback

Aural and visual feedback had the lowest mean absolute pitch error in both the common repertoire and unstructured time sections even though we saw evidence in Section 9.8.2 that aural feedback takes time to get used to. Similarly, participants overwhelmingly selected it as their hypothetical preference for practice (8 of 12). As demonstrated in Table 9.9, and in Section 9.6.2 participants valued the ability to switch between feedback options depending on needs (B2, B4, B5, B8, B10, B11, B12). Aural feedback presented a useful alternative to participants preferring visual feedback when playing with a score or unable to devote sufficient visual attention. Participants, hearing error identified by aural feedback, were also able to use visual feedback to clarify if unsure how to correct it. B11 and B12 both referenced they appreciated being able to look at visual feedback to confirm notes heard through aural feedback were correct and in tune. B11 said:

> I would use both, sometimes it's nice to see if I'm hitting a particular note (such as $C\sharp$). It's nice to have the visual on, but in terms of actual playing I would use the headphone and my ear and then every now and then I'd look over and see if I was getting a green on the $C\sharp$. It's really good to have both, it was nicer than not having [visuals] at all. (B11)

Aural and visual feedback will suffer from the same familiarity problems that both aural and visual feedbacks have. It will not fix flicker in visual feedback, can clarify confusion from aural feedback but will not fix potential overstimulation but having both does give the participant the option to switch focus if desired. There was some evidence that trying to use both feedback methods simultaneously was overwhelming (B2, B11). For instance, after attempting to use both aural and visual feedback simultaneously, B11 commented, "[Correcting pitch] is much easier without looking" and stated she subsequently stopped using visual feedback regularly except to confirmation correctness.

9.8.5 'Double Headphones'

Following on with our investigations into complexity management and simplifying the violin, participant reaction to our simplified case using double headphones was a pleasant surprise. Based on the negative reaction by experts to full pitch snapping (Chapter 8), we were not sure whether participants would like the experience, and moreover, whether they would find it useful beyond just a gimick. With 10 out 12 participants liking it, and the two remaining both calling it fun at times, support for the simplified experience was strong. Not only did younger students enjoy it, but so did the adults.

We had also not expected participants to articulate that they thought wearing both headphone ears would be useful for learning. Within the study, students and parents in particular might ask why we would use one of the main forms of feedback, but as the double headphones section was only the last few minutes at the end of the lesson, the author never had to say much about why it was included. As a result, there is a reduced chance of bias as participants were unlikely to have derived meaning for double headphones from the author and all ideas about its use came directly from participants.

A surprise result was that participants did not seem to notice the difference between half-snap and full-snap cases even though we found statistical support for the difference between the two cases in the study with experts. When asked if what they though they sounded in tune using the double headphones most students said yes (B2, B4, B5, B6, B10, B11, B12). Similarly, numerical analysis of pitch error between the two cases found no clear difference. There was less than 1% difference between the full-snap and half-snap performance.

Both the perceived and the numerical results may be due to a less experienced ear. Only two participants appeared to noticeably improve (2 cents difference) when using half-snap: B5 the most experienced violinist in the study, and B11, the professional percussionist. B5 performed 13% and B11 performed 17% better in half-snap cases which is on par with the differences seen in the pilot study. Though intonation expertise may explain why participants said the half-snapped double headphones were in tune, responses may also be influenced by trust and acquiescence bias. Participants were unaware of the use of half-snap so may have assumed they were wrong about perceived pitch error and that the 'correct' answer when asked if audio was in tune was 'yes'.

An Intonation Aid, Not Just Motivation Aid

Even though we had intended double headphones as an enjoyable low stress opportunity for sounding in tune in order to help motivate and engage students, as suggested by comments in Section 9.6.2, study participants suggested playing with both headphone ears was potentially helpful for correcting pitch and liked it as an intonation aid. Comments mention how using both headphone ears forced participants to block out external activity and fully focus on what they were hearing. Further some users commented it helpfully highlighted error. For instance, while it might expected for a student to notice when they are repeatedly playing a note dramatically out of tune, either with but especially without additional feedback, they often fail to fix the error. However a severely out of tune note may snap to the wrong note successfully drawing their attention to the error. Alternatively, if a student is not sure of the correct note they can guess and the snapped version can help clarify whether it was correct or not regardless of whether it was in tune (see B12 Section 9.6.2).

How wearing both headphone ears more successfully highlighted more subtle intonation error is not entirely clear. Discussing why he liked the double headphones for helping with intonation, B2 had difficulty describing what the effect was saying,

> I was gonna say I found... when I'm a bit off, it sounds different... it definitely sounds a bit off when you're not on the right [pitch].

Similar to the expert study, if a student did not use sufficient pressure or their note was exactly between two notes, audio would glitch which might be interpreted as a note needing correction. However, in some cases the audio in the headphones did not glitch. Potential highlighting of error relates to audio bleed and the statement from Section 9.8.2 by B5 that during single eared headphone use when playing in tune the aural reference and the performed audio blend whereas out of tune notes will clash highlighting error. How much the participant can hear their own playing is dependent on headphone volume and some students did report being able to hear both (B1, B5). However B2 suggested he could not hear the acoustic violin yet could still recognize error. Regardless of the mechanism, B2 indeed correctly fixed notes while wearing both headphones.

'Let Yourself Just Play'

One of the ideas behind the use of double headphones was that it would allow the user to just play. For a beginner challenged by playing the correct notes, playing them in tune, and also maintaining reasonable tone double headphones was hoped to provide a chance to perform the piece without the usual mental stress, concentration, and skill required to play well. B5 discuses this in Section 9.6.2, saying essentially that if you play the right notes, regardless of intonation, you have the chance to hear the "Piece played perfectly" by you. B5 captured the spirit well.

Further, it makes it easer for a beginner to play a song they do not know well. In the two instances where a participant played a piece using double headphones that they barely knew, B7, who did not know either common repertoire piece, and B11, who asked to try the double headphones with a piece she had just started, the snapped audio sounded reasonably passable and identifiable in comparison to the actual performance which sounded much worse. As B11 stated:

> We always did double headphones at the end of a lesson when I'd already gotten better, and that was fun anyway, but actually when I start the lesson, or even better, a new piece I think playing with both headphones would make it much more enjoyable, less stressful, less things to think about, and I could get my fingerings together. And then I'd take one headphone off, and I could focus on the actual pitching, and I'd already be relaxed and I'd know where I was aiming for. That'd be cool.

Eliminating Problems from Acoustic Bleed

After the first study, acoustic bleed was a major concern. Though it is not conclusive that that was why experts judged pitch corrected audio to be more out of tune than uncorrected audio, audio bleed was a major suspect. As discussed in Section 9.2.1 for the study with beginners we added compression to pitch corrected audio to help participants distinguish between corrected and uncorrected audio and also used acoustic instead of electric violins.

In the study with beginners, there were no complaints of pitch corrected audio being out of tune[6]. All participants who responded to questions whether corrected audio sounded in tune (B1, B2, B4, B5, B8, B11) stated that playing with both headphone ears sounded more in tune than when they played normally. Between the 100% response rate that the audio sounded more in tune and the overall positive response to playing with double headphones, it suggests were successful in removing the negative experiential effects of audio bleed.

9.8.6 Trust and Authority

Trust is crucial for pitch feedback tools to be useful. If a practice aid is frequently wrong, it loses its value. Students appeared to generally trust the augmented violin feedback with B4 stating a reason he liked the feedback methods was because, "They give you true feedback, they don't always give you positive." As the author was an expert and the beginners were working with technology, there was a high degree of inherent trust that the technology the author/teacher was offering was trustworthy. This was true even when it struggled with the three lowest notes on the violin. Asking B6, who has perfect pitch about visual feedback at the bottom of his scale:

> A: "Do you think it is right at the end?"
>
> B6: "I think it has to be, Laurel."
>
> A: "Actually no, at the bottom it gets wrong and you get right."

The interaction with B6 was typical and clearly demonstrated the high degree of trust most students had. Trust and authority were necessary to encourage participants to use feedback, but they had the drawback that participants were less likely to report issues during the experiment. For instance, with aural feedback, if the volume in the headphones was too high, rather than speaking up participants would often tolerate it. As mentioned in Section 9.8.2, B8 spent 26 minutes in one lesson before asking to take off the headphones. It turned out a large issue was the audio was too loud, but by this time, he was also feeling very negatively towards aural feedback while in the other lesson including aural feedback he had reacted positively to it. Similarly, if the volume was too low and the participant could

[6] As will be discussed in Section 9.8.6, there was a software flaw during the first half of the study. Two students who reported hearing out of tune audio before the software was fixed stated it sounded in tune after.

not effectively hear the guide pitch, the feedback became unusable. Only the adult participants ever asked for volume to be turned up.

B8 is also one of the few participants to express any distrust of the system (Section 9.8.2). B1 also suggested the augmented violin was wrong stating, "It helped when I got the [second fingers] a bit off, but when I did the song just now, sometimes it just played notes, random notes, when I played." As listed in Table 9.9, B1 stated his favorite lesson was having no feedback.

Trust and authority roles also meant that incorrect pitch settings or sensor issues were also slow to be caught. There were multiple times where the author initially failed to correctly set the scale for a piece. Part of why lessons with aural feedback almost always included a small speaker was to help the author catch such errors, but only B4 and B11 ever notified the author that a setting was incorrect. Further, while sensor functionality was reliable, connectors sometimes slipped, degrading performance dramatically. Again, these issues were often slow to be discovered as, apart from B1, B4, and B12, participants did not identify incorrectly tuned notes, or excessive audio artifact.

As was mentioned in Section 9.5.1, during the first half of lessons there was a flaw in the hardware and software design that allowed the whole tone above the open string to often be estimated roughly 70 cents flat. It is difficult to know how much the issue affected participants' experience in part because it affected different people differently. For instance, in the worst case, looking at someone with insuficient finger pressure, calculating error using the original flawed pitch estimates in the lesson increased error 74%. Whereas in another lesson with a student with stronger finger pressure it only increased error 1%. Although no student complained the aural guide was incorrect, when asked if they were in tune after the double headphones section both B5 and B8 stated it was out of tune. After the software was fixed, when asked the same question after the double headphones section both participants stated it sounded in tune. This suggests the error may have been noticeable when listening. Errors in the pitch estimate will undermine trust and usefulness. Indeed, the author also had the impression that after the change in software, aural feedback was better received.

9.8.7 Tool or Toy?

An important question to ask especially as qualitative results often counter quantitative results is whether positive responses were due to novelty. Most of the participants were excited to play with double headphones, but was this just because it produced interesting sounds? Due to the time and resource limited nature of the study, we can not conclusively claim participant enthusiasm would persist with more exposure. However there are reasons to believe interest would be sustained over time. Although not intentional, Question 1 from List 8 asking about a student's favorite lesson, often ended up yielding which single feedback method, visual feedback or aural feedback, participants enjoyed more. Meanwhile, Question 4 from List 8 asking what feedback method a student thought most helpful during practice, was more likely to yield an answer based on which feedback method participants found the most helpful. Answers to Question 4 were more likely to include statements that demonstrated an understanding of why and when a feedback method was useful suggested participant response may have been less novelty based.

For instance, B8 reported, "I like [the visuals] because I try and make it as green as possible. I don't like it when it's red, because I don't like to make a mistake on the violin. It kinda annoys me. Plus green is my favourite colour." Liking the color green is a trival response, however B8 also showed interest in attempting to use the visual feedback effectively to minimize mistakes on the violin. Further, B8 expressed interest in using the displayed pitch names depicted to help learn notes (see Section 9.8.3).

Additionally, which feedback modality a participant liked most did not necessarily correspond to which they thought was most helpful. B6 loved using the headphones, and especially the double headphones at the end of a lesson. However B6 would often appear to experiment with the aural feedback, playing songs in the wrong key. In two different lessons, he played the common repertoire song Long Long Ago in F minor before agreeing to play it in the correct key. As the teacher, the author could not find any indication he was actually using the feedback to inform his pitch. Still, while use of aural feedback appeared to be mostly fun, B6 engaged with the visuals more constructively. When asked what he would find most helpful for practice, he reported he found visual feedback the most informational suggesting B6 was able to separate which interaction for him was more of a toy and which was more of a tool.

9.8.8 What People Did Not Like

Though response was largely positive, we wanted to give participants the explicit opportunity to be negative about the augmented violin and the pitch feedback methods. In Question 3 of List 8 we asked what people liked least about their experience with the augmented violin. A common complaint was excessive and bulky cabling (B3, B5, B9, B11). Cabling is indeed bulky in the current build, but not inherent in the augmented violin design and could be addressed with improved manufacturing. Two people expressed that they did not like the actual violin. This response is not surprising and justifies our goal of cheap, reversable augmentations so people can add augmenting sensors to their own instruments, but the use of a study violin was a necessary practical constraint at the time of this study.

The one complaint likely to have significantly affected user reaction was audio burbling and clicking. For instance, in B1's last lesson, he reported "[The audio] just had random clicks all the time." and, having previously responded positively to headphone sections, became very negative about any use of the headphones. Audio burble and clicking had multiple sources. As discussed in Section 8.4.7, users with poor finger pressure (B1, B7, B8), are likely to hear burble. The risk of burble from low finger pressure is inherent in the sensor informed pitch estimation, but, with the more tolerant revised pitch estimation algorithm used in the second half of the study, one we consider acceptable. Good finger pressure is part of Western violin technique. Additionally, burble occurred when sensor connections became loose, an addressable manufacturing issue.

Lastly, clicking and pops could occur due to host computer issues, not the augmented violin itself. The augmented violin software, logging and the recording of audio data for the study required substantial processing power. If an unrelated background process on the computer started demanding memory and CPU, the audio host would miss audio deadlines. Processor strain could have multiple effects, most commonly the occasional click, progressing to heavy clicking and delay or worse. The same morning as B1 complained of clicks, B4 remarked on a particularly distracting episode,"With the headphones there were sometimes ... when it just glitched, and I had the headphones on one time and there was a huge buzzing then the computer went crazy." In B4's case, the severity of clicking and popping clearly caused him to lose concentration. Though no one else reported similar severe audio glitches impacting concentration, the author observed a couple instances where headphone noise occurred very close to participants' losing focus. Shutting down other programs on the computer typically solved the problem suggesting this issue could be solved by giving the augmented violin its own dedicated processor.

9.8.9 Persistance Effects

It is important to point out that in this study, we did not test test whether benefits from using either form of feedback were retained once removed. Persistence effects are more of a concern with visual feedback. O'Connor, Van Der Linden, and Percival [124, 169, 136] all express concern about making decisions based on a correction modality not normally available during performance, with Percival skeptical of real-time visual feedback in particular. This was commented on by one of the study participants.

Though B11 found visual feedback often easier to use than aural feedback, B11 similarly recognized that relying on the visual feedback to determine how to correct pitch was not necessarily the most helpful feedback modality in the long run,

> I think... the reason probably I prefer this is that the hearing thing stresses me out because my ear isn't used to this and part of what I wanted to get out of doing some violin lessons was actually to improve my ear so maybe actually what I need to do is more stuff with the headphone. That'd be useful for me.

Still, in Table 9.9 B5 points out that in some cases, visual feedback can be uniquely helpful illuminating error that is hard to hear, and up to a point, the more work a student does on fine tuning work, the better their ability to differentiate correct pitch. It would be necessary to design a follow up study to test for sustained benefit from either mode of feedback.

9.8.10 Polite versus Impolite Feedback

One of the interesting issues that arose was the difference between polite means of highlighting error versus impolite aggressive ways of highlighting error. Though we attempted to ensure aural feedback sounded smooth and did not glitch, there is evidence that that might not necessarily be optimal. For instance, B11 commented, "Yes [aural feedback] is useful, because I can hear it being [the correct note]. Because I can hear it right, it almost means I don't correct." If a student does not correct, the aid becomes counter-productive. However, we found situations where the unpleasantness of audio glitching caused strong corrective reaction in users.

Due to unusually low audio input levels, there were two days where audio would glitch in relation to how out of tune the player was: the more out of tune, the more noticeable the artifact[7]. B2 said, "It's almost irritating, you wanted to do something about it...if I played the wrong note it [was like it] gave me an electric shock." B2 subsequently repeated the aural feedback lesson with the input levels fixed and stated:

> [The normal sound] does help, it's more clear when you play the wrong note without being annoying, because [before] it was like, oohhh, but, I don't know which one is better; probably the annoying part is good because then it forces you to not miss that note.

It would be interesting to investigate whether impolite highlighting of error would be more effective at helping students improve or whether the unpleasantness would instead lead to students not wanting to use the aid at all.

[7]The retuning VST only snaps audio over a certain volume. When the level of the audio input signal was sufficiently low, pitch correction would turn off and on depending on instantaneous changes in volume. As a result, the greater the difference between the played pitch and the target pitch, the more noticeable the audio artifact became.

9.8.11 Compression in Aural Feedback Audio

As we stated in Section 9.2.1, we added heavy dynamic range compression to the pitch-corrected audio to make it easily differentiable from the performer's acoustically produced audio. It turned out the addition of compression yielded an interesting useful discovery: it largely flattened tone quality. Good versus poor tone quality was essentially indistinguishable in the compressed audio. Only the worst tone, acoustically generating harmonics or being scratchy, degraded the compressed audio quality. Asking B11 if she was able to hear audio variations due to poor bow technique, she replied, "No. Unless sometimes I was on the wrong string obviously, but apart from that, no I couldn't hear any difference." The use of compression does decrease the expressivity of the instrument, flattening dynamic changes, but this flattening in tone is useful as it meant that when using both headphones, users could relax not just pitch accuracy but bow technique without penalty.

One of the main goals in this thesis was to look for ways to make the violin learning process easier and more enjoyable. The ability to remove all but the worst effects of bad tone is a useful result for future experimentation. Compression of the acoustic audio does alter the violin sound, but participants generally suggested it was fine, complaining far less than users of the electric violin in Chapter 8. In the future it could be added as an intentional variable in learning with the augmented violin, adding or removing compression depending on where the student wants to place focus; pitch snapping with low compression might be good for focusing on bow work while no pitch snapping with compression might be good for focusing on pitch without worrying about the bow. Additionally, for future experiments into complexity management, we now have a way to simplify the audio impacts of bowing technique.

9.8.12 Success as General Use Real-World Tools

A major target of our augmented violin was that it was usable in a real-world non-laboratory environment. Based on this, our requirements were that the augmented violin should be real-time, low-cost, non-intrusive, portable, robust, easy-to-use, and that it should not encourage bad technique. Adding to the results of the pilot study with experts, the study with beginners provided further support for out augmented violin being effective in real-time use and that it was non-intrusive; no participant reported any distracting audio or visual delay, nor any issues with the technological addition to the violins apart from the extra cabling. Of the two adult users with the augmented bow, neither had issues using it or reflected negatively on it.

Robustness, portability, and technical ease-of-use are demonstrated simply by the practicalities of the study itself. 53 lessons took place on 20 days in three different locations eight miles apart. The author commuted to lessons carrying all augmented violin and study tools including video camera on bicycle and did not require any additional equipment at the destination beyond a camera stand. Neither of the violins, nor the augmented bow used during the study suffered significant technical issues. Removable connectors did come loose on occasion, but that was easily fixed in the moment and can be easily addressed in future construction. Further, at St. Matthew's Academy, due to limited access, we had only had a 10 minute time for setup. 10 minutes proved sufficient for setting up, calibrating, and testing the augmented violin, along with preparing the rest of the study environment. Though additional simplification and improvements of the augmented violin would be necessary for general release, the augmented violin and associated tools proved themselves in a real-world test case.

Our last requirement, determining whether the augmented violin enables or encourages bad technique would require a longer term study, but during basic play, as a teacher, the author did not see any indication the augmented violin itself was negatively affecting technique. Of the two using the augmented bow, neither B2 nor B11 had any problems with the augmented bow and stated using it was fine.

9.9 Reflections from Teacher's Perspective

As the teacher conducting lessons and watching the use and reactions of students to the different feedback methods, my impressions, though likely biased, remain relevant and insightful. Overall, students were very positive towards all three methods of feedback, aural only, visual only, and both combined. Just as Johnson found in her studies [74, 75], I found that feedback preferences varied between students. Personally, even as a professional, there are times during practice when I would appreciate an aural guide snapped to certain notes and also having more useable visual feedback than the tuner to check specific notes.

9.9.1 Perceived Reaction and Use of Feedback Aids

Students liked aural feedback for the reasons we expected: it helpfully highlighted error, giving the correct version that most students could intuitively follow once they got used to it. As a teacher it was hard for me to tell how much students were using it, leaving me sometimes surprised that some students did not react. It is also true that once the algorithm for pitch detection had been expanded to capture first fingers more easily, reaction to aural feedback seemed more positive with participants and it was easier to get them to discuss it with me.

Although it is not evident in the numerical data or student quotes, from my teacher's perspective it appeared that very few students actually used the visual feedback. Even though we annotated gaze, suggesting graphics were used 16% of the time, I did not think students were often responding to it. Both adults, B2 and especially B11, seemed to make a continuous effort to look at and respond to the visual feedback, but otherwise, despite comments saying they liked it, only B3, B6 and B8 showed any extended interest beyond one or two scales or brief experimentation after I discussed how to use it. Some students even appeared to completely ignore it (B9). Apart from B11, and possibly B4 who said he kept the visual feedback in his periphery, I do not believe anyone used visual feedback while performing the common repertoire piece.

As a result, though I have treated it separately in the numerical analysis, my subjective impression was that lessons with visual feedback were effectively the same as lessons with no method of feedback. That is not to say visual feedback is not potentially a useful tool, only that in the lesson context with me present, I do not believe students used visual feedback significantly.

Students were particularly excited about the double headphone section of the lesson, often smiling or exclaiming when I announced it, or asking if we were going to do it earlier in the lesson. Even students like B1, who disliked and distrusted the aural feedback smiled saying he enjoyed using the double headphones. Some of the students less enthusiastic about double headphones made comments to suggest that they were frustrated because they were trying to play well normally and found it hard when they could not hear what they were actually doing. Others, like B5, recognizing he could not hear himself, just decided to have fun.

9.9.2 Teaching with Feedback Aids

As a teacher, I found both methods of feedback useful as teaching aids, visual feedback more so than aural. Aural feedback was effective as a concept but less useful as a teacher in that if aural feedback was unavailable, I could easily play with the student to achieve similar effect. Visual feedback however was useful in pointing out major error or helping the student if I asked "What's the name of the note you are you playing?"

Though overall I felt visual feedback was a useful teaching aid, there were times it was more of a distraction. While giving oral instructions, with some students I was worried about losing their focus to the visual graphic (B6). One student, who did not complete the study, came across as particularly distractible and even though I wanted to reference the visuals, I did not as I was too worried I would lose his focus and would have to start the topic I was teaching over again. Another time, B3, while playing a piece by memory for me during unstructured time was occasionally getting lost. Her first instinct would often be to look at the visual feedback which rather than helping seemed to make her more confused and less likely to get back on track.

Aural feedback was also sometimes confusing if I failed to set the key correctly. I would sometimes only notice my error due to students repeatedly playing closer to the wrong accidental (B1, B8, B9, B10). It was also clear that when there were issues with the augmented violin technology it became unhelpfully distracting. For example, in a lesson with B9, I heard a significant glitch in the audio and noticed B9 appeared to lose concentration immediately after. But apart from the occasional technical issue not surprising in a prototype, the aural feedback seemed more help than hinderance.

In contrast, one benefit of aural feedback was that though it highlighted error, it did so in a non-judgemental way. Percival [136] points out that teachers filter error discussion with students for both musical and psychological reasons. If a student receives too much negative feedback it is discouraging. Although no one admitted to it, I got the distinct impression some students (B3, B10) got discouraged while looking at the visual feedback. Out of tune notes were orange or red and, having previously been confident, after seeing lots of red while playing some students' demeanor changed. They became more subdued. I did not notice students clearly discouraged by aural feedback. The visual feedback could be softened to a degree by altering the colors so that error is less aggressively highlighted.

9.9.3 Practice versus Lessons

One final reflection is that though we tested in a lesson context, these feedback methods were designed with practice, not lessons in mind. Aural feedback was not particularly necessary in the lesson as I could just as easily provide an aural guide by playing with a student with the added benefit that they could watch my fingers. Students often expected me to play with them which, as B11 pointed out in Section 9.7.1, actually undermined the aural feedback. However at home, I am not there to play with them, but aural feedback could serve a similar purpose.

Additionally, I was regularly challenging students with hard tasks and would judge the result. Pressure to immediately achieve tasks meant students would often do what was expedient, and it was often risky to take the time to focus on the feedback tools. Practice can be much more experimental.

Time was a major constraint and at a premium when conducting the study as lessons. Two 30 minute lessons with each feedback method is too short to really test the impacts of pitch feedback. I believe a better judge of our feedback methods would be to build additional instruments and distribute them for a similar length study, but used during practice.

Johnson [76, p.56-58, 67] also discusses that feedback needs to be appropriate to tasks a student is working on. Roughly 15% of playing time in lessons was spent explicitly focusing on tasks where attempting to play the right note in tune was a distraction from what I was asking a student to concentrate on, yet for simplicity and keeping lesson interruptions to a minimum students were still provided with feedback. With either longer lessons, or asking students to use in home practice we could more easily use intonation feedback aids only when helpful for the task being worked on.

An Aside

An interesting side note demonstrating how hearing pitch is not inherently natural occurred when we demonstrated the augmented violin as part of 2016's Sonar+D in Barcelona. As we've said, good pitch is not just putting the finger in the right place, but being able to hear the right pitch and know if what you are playing matches it. During a public demo, passers by were invited to try out the pitch snapped electric augmented violin from Section 5.7.1 without explanation. Members of the public who had not played stringed instruments universally expressed no understanding of what was happening: that every note they played was somehow well tuned. Even turning snap on and off while they were playing rarely elicited a response. It was only when the author demonstrated a glissando with no snap and then a glissando with snap that people would show the glimmer of understanding. In comparison, public who had significant string training almost always immediately reacted, expressing how odd it felt losing fine control of their pitch.

9.10 Conclusions

We ran a four lesson in-situ study with 12 participants in order to investigate the potential for real-time technology to act as an intonation practice aid for violin. Participants tested four different types of pitch feedback: 1) aural feedback in the form of a single headphone ear providing a guide pitch created by processing the audio and correcting the user's playing to the nearest note in the scale, 2) visual feedback in the form of a graphic displaying the name of the note the user is playing and colored bars depicting level of intonation error, 3) combined feedback providing both the aural and visual feedback, and 4) no feedback beyond the violin itself. Additionally, we asked participants to play using both headphone ears. This was a follow-up to the study with experts in Chapter 8 where we tested the effects of artificially correcting intonation error. as follow-up to the expert study testing the effects and experience of reducing pitch feedback.

Participants generally responded positively to all types of feedback even though statistical analysis does not show any clear effect of increased pitch accuracy. 7 out of 12 participants responded saying that lessons with aural feedback were their favorite even though aural feedback resulted in worst pitch accuracy. 8 of 12 students thought combined aural and visual feedback would be the most helpful for individual practice. Both aural and visual feedback were praised for highlighting error with aural feedback more helpful at providing the correct target, more relevant to how a student self-corrects, while visual feedback was praised for being easier to identify how to correct. Only one student said they preferred having no feedback and only two students said they were unlikely to use our intonation aids during practice.

Visual feedback often suffered from the need for visual attention the need to compete for visual attention. How exactly (and potentially incorrectly) aural feedback was correcting was also sometimes unclear and also took some time to get used to, an issue that lead to our largest changes in pitch error. Having both aural and visual feedback was praised for allowing use of aural feedback when visual attention was needed for other tasks while also having visual feedback available to clarify information from aural feedback.

In contrast to our pilot study, wearing both headphone ears and hearing pitch corrected audio ('double headphones'), was widely enjoyed. 10 out of 12 participants stated they enjoyed the experience and would be interested in being able to use it in future. Even the two that did not express repeated interest in the experience said they enjoyed it at times. Three students remarked they enjoyed it as it effectively meant they "Sounded awesome" (B5) without much effort. Though issues likely due to acoustic bleed seen in the pilot study seem to have been largely eliminated, 3 students expressed that they though the double headphone experience would be most helpful for practice. Students appreciated that it assisted focus and forced them to listen to what they were doing.

We included two case studies (Section 9.7), one detailing focused use of all feedback methods by the participant resulting in an experience closer to what would have been asked during a laboratory style test, and one demonstrating the effects of a participant's use of aural feedback. We also discussed issues relating to participant bias when using and answering questions related to feedback methods (Section 9.8.1), how students largely trusted the intonation aids (Section 9.8.6), and looked at the possibility but unlikelihood that positive remarks were because the experience was novel rather than valuable (Section 9.8.7). Negative response to the system tended to center on excessive cabling and audio glitching (Section 9.8.8).

We also had two unexpected findings interesting for future investigation. The first was that when we accidentally aggressively highlighted intonation error by setting input audio levels too low so that pitch error would increase audio glitching, students felt compelled to try and fix their intonation (Section 9.8.10). Otherwise, when using the aural guide they might notice they were out of tune, but not feel sufficiently motivated to fix it. Additionally, we used heavy compression of audio in order to differentiate the sound of snapped audio from the acoustic sound generated by the students' playing. The compression process effectively flattened bow technique so that only the worst bow technique was differentiable from good bow technique (Section 9.8.11). As we are interested in simplifying violin playing for experiments into complexity management, being able to remove error in audio from bow technique allows us to experiment with that parameter in future, including investigations into whether there are benefits to simplifying demands for good bow technique while working on pitch, or simplifying demands for correct intonation while working on bow technique.

Additional future investigation would be worthwhile exploring our intonation aids in home practice. Our study was conducted in the form of lessons for practical reasons even though the intonation aids were designed with practice in mind.

Lastly, this study proved basic robustness, reliability, and usability of the augmented violin (Section 9.8.12). 53 lessons took place involving 14 people and three non-specialized locations. The augmented violin and associated study tools were cycled between locations and suffered only minor technical issues even when set up time was limited. The augmented violin proved itself well.

Chapter 10

Conclusions

10.1 Review of Goals and Achievements

We set out in this thesis to investigate the use of technology to assist violin learning. We focused on two main learning interactions: practice aids for improving intonation, and whether we could improve learning efficiency through the use of complexity management, intentionally altering the inherent difficulty of an instrument in order to assist practice motivation. Due to the lack of an augmented violin or tracking system appropriate for tracking violin performance actions in a non-laboratory environment, we set out to design an augmented violin for use in real-world educational contexts. Based on pedagogy suggesting real-time interactions are better than reflective interactions [85], we opted for our augmented violin to be real time, along with practical requirements for it to be low cost, durable, portable, non-intrusive, and instrument safe. As we are focused on learning, the augmented violin must also play like a normal violin.

10.1.1 Our Augmented Violin

We accomplished the requirement for a practical real-time instrument by designing an appropriate augmented violin and bow. Augmentations to the violin, described in Chapter 5, consist of the addition of a fingerboard sensor for tracking left-hand finger position. Attached to the fingerboard like a sticker, the fingerboard sensor is effectively four custom-made linear potentiometers: one per string. These provide a rough estimate of pitch played useful for low latency pitch estimation and event detection, as demonstrated through a case study of note onset detection (Section 6). Combining the rough hardware pitch estimate with established audio pitch estimation techniques allows high-accuracy pitch estimation even at low latencies.

The augmented bow, presented in Chapter 4, uses a novel approach to bow tracking by using optical sensors to measure bow hair displacement at different locations along the bow. A mathematical model of the hair displacement for different bow positions and pressures is created by training for a given bow with sensors attached. The mathematical model can then be used to estimate bow position and pressure based on sensor measurements of hair displacement only. Though our augmented bow does not match the potential accuracy of video motion capture or EMF techniques, it works sufficiently well for pedagogical purposes with an average normalised RMSE error of 6.3% when measuring bow position, excluding when playing very close to the frog, which is highly unusual in normal performance. It also performs well estimating pressure, accurately detecting major bow force actions such as accents and setting the bow on the string. Uniquely, the entire augmented violin, including both bow and fingerboard sensors, is also low cost and portable with everything needed to use it (apart from a computer) costing less than $150 in materials, easily 10 or even 100 times less than the higher accuracy techniques.

Proof that the implementation is robust, portable, and reasonably easy to use was provided during the extended study conducted in Chapter 9. The study featured 14 different players taught in 53 lessons on 23 days over three months and took place in three locations up to eight miles apart. All tools were transported without a car and set up in limited time. Despite heavy handling test instruments had no significant technical issues and remained in use after study completion.

10.1.2 Intonation Aids with Beginners

The second target was to investigate how the augmented violin could be used to aid intonation through corrective feedback (Chapter 9). We chose to look at three potential pitch feedback options. One of the options mimicked the traditional practice of playing with the teacher as a guide; we gave students a pitch corrected version of their performance for listening to in one ear, while they were able to hear their violin normally through the other. The second was an attempt at making practice with a tuner, a visual feedback tool already used when practicing careful slow intonation, more practical by designing a faster, more responsive version. The third was providing both aural and visual feedback in the event that the combination was better than either individually.

In a real-world study with 12 students taking four lessons each, we found positive response and performance impacts for both the aural and visual feedback enabled by the augmented violin. Some participants found aural feedback very useful for reducing the mental requirements to detect incorrect intonation and that the audio of the corrected pitch provided a useful reference tone for guiding their own pitch correcting actions. Similarly, many users thought visual feedback was also clear for highlighting error and even easier for interpreting how to correct pitch than the aural feedback method, but required more attention. Students predominantly (8 out of 12) thought the combination of both modalities would be the most helpful during practice. Though not statistically significant, examination of as-played pitch during lessons with the different feedback methods backed up some students' self-reported reactions, though overall suggested no feedback was best. As van der Linden and Johnson suggested would be likely in non-laboratory studies [168, 76], we found performance using the different feedback methods was highly diverse, impacted by different learning preferences, individual student focus on the day, and lesson tasks.

An unsurprising result was that many participants reported that visual feedback was of limited use when performing tasks that otherwise demanded visual attention. For instance, when reading a score, participants had to look elsewhere, but when performing a piece from memory, it became quite helpful. Additionally, though the update rate of the visuals was fast enough that users could monitor their pitch error during most performance tasks, the speed also meant many participants complained the visuals changed too quickly so that they became more of a distraction than a help.

In contrast, aural feedback was always passively available but required balancing aural focus between two sources. A surprising result was the length of time it took users to acclimatise to the addition of a second audio source and the importance of correctly setting volume to make the pitch corrected version easily distinguishable from their non-pitch shifted performance. Similarly if the shifted audio was too close in sound to the original, it was confusing which sound the user was controlling. One of the strongest numerical trends from pitch analysis of playing was that the addition of a second audio source initially led to poorer performance, though with time, the negative effect would disappear, sometimes replaced by a positive effect. This result suggests that to properly evaluate the helpfulness of aural feedback, we should run a longer study.

10.1.3 Exploring Pitch Simplification

Motivated by the idea of optimizing an instrument's complexity, not just providing a practice aid, we started investigations into simplifying the violin and what effects that might have on performance and player enjoyment. Intonation is one of the most important yet frustrating tasks impacting performance quality. We theorize simplifying pitch may lead to earlier enjoyment of violin achievement and this thesis marks a beginning step into investigating whether this might be true. Within this thesis, pitch simplification was performed by having study participants wear headphones on both ears in order to block out the acoustic sound of their playing and replacing it with pitch corrected versions. We looked at the performance and reactions of participants to different levels and styles of pitch correction (Chapter 8 with follow up included in Chapter 9).

Response to pitch simplification was wide-ranging but largely tied to experience. Experiments were carried out across two studies, and incorporated a broad variety of participants, both young and old, with a wide level of expertise, from 30 years experience (Chapter 8) to 1 year of experience (Chapter 9). In the initial in-depth study with experts (Chapter 8), we confirmed that correcting heard pitch and thus removing feedback for intonation around a specific note resulted in worse pitch performance. Restoring partial feedback resulted in significant improvement, though still worse than the control case without any pitch correction.

However, in a follow up included in the study of intonation aids with beginners (Chapter 9), there was no clear difference in performance between fully pitch corrected cases where intonation error was eliminated in the headphone audio or partially pitch corrected cases, where the player could hear a reduced version of their intonation error. Further, and somewhat surprisingly, with beginners, neither pitch corrected case was clearly worse than playing the violin without pitch correction. The difference in results between the two studies is likely a reflection of differences in the level of learned pitch differentiation between beginners and experts.

User response to reduced feedback modes between the studies was also drastically different and dependent on experience. Experts were generally uncomfortable with the loss of full pitch control while beginners enjoyed the reduction in error. The inverse link between expertise and positive experience was first suggested in the expert study (Chapter 8) and strongly supported in the follow up study with beginners where, rather than discomfort, there was almost universal enthusiasm for the corrected pitch experience (Chapter 9). Additionally promising is that though experts disliked full pitch correction, they were not clearly negative about partially pitch corrected audio suggesting that what is important is to retain some ability to influence the output and that instrument play retained the same basic rules.

The two studies took place on two different violin designs (Chapter 7), first electric then amplified acoustic with heavy compression, so some of the additional enthusiasm might be attributable to the changes in audio. Indeed, we suspect perceptions from the first study, conducted with experts using the electric violin, may be due to hearing both corrected and acoustic audio but not being able to distinguish between the two. We altered the audio based on this suspicion for the study with beginners using the acoustic violin, but the magnitude of the switch in enthusiasm suggests the change in violins and bleed is at best a contributing factor. Results with beginners suggest the reduction in complexity was not only fun, but useful, though users suggested they needed more time to get used to it. It is only with a significantly longer longitudinal study that we will be able to evaluate whether our attempts to reduce the barriers for a beginner to achieve musical success have the hoped for motivational effect.

10.2 Review of Contributions

Here we review the list of contributions presented in this thesis:

Design and use of optical sensors for low cost, low latency real-time bow tracking (Chapter 4): we introduced a new method for bow tracking, novel both in technology and theory, that may be useful for others similarly searching for low cost practical means for bow tracking. Besides use in teaching and practice we demonstrated *using our augmented bow for detecting real-time note onset* in stringed instruments (Section 6). Implementation details are expected to provide digital instrument designers on the practicalities of optical tracking, along with bow tracking specific details.

Low latency real-time pitch and note estimation for the violin through fusion of sensor and audio analysis (Chapter 5): though audio-only means of pitch detection have reached reasonably high accuracy, performance suffers in a real-time, low latency context. We used custom linear potentiometers to detect finger placements on the fingerboard and used the resulting rough hardware based pitch estimates in conjunction with audio estimation for effective low latency pitch tracking. We were also able to *identify note onsets occurring due to fingering changes in real time.* Though we are not the first to put linear potentiometers on a fingerboard, ours were heavily tested to work reliably and rated virtually unnoticeable and highlight the simplicity and effectiveness of our home-made linear potentiometers.

Demonstrated effects of altering levels of pitch feedback on un-fretted performance and the link between user skill level and perceived experience (Chapter 8, Chapter 9): we validated that heard pitch is a key part of a violinist's ability to play correctly in tune even when playing simple tunes. Though this may seem obvious, we have not seen anyone verify this particular aspect of intonation execution. Furthermore, we demonstrated that partially restoring heard pitch error improved performed pitch execution beyond the level of actual error heard. Our studies also suggest that correction of heard pitch impacts beginner violinists less, suggesting they have not yet developed a strong internal pitch feedback loop. All of these issues are relevant to pedagogical studies of non-discrete pitched instruments and support traditional yet unresearched teaching practices. For our own purposes, it provides a baseline starting point for whether pitch simplification is a useful area for introducing and testing complexity management.

Demonstrated, through a real-world study with beginners, the potential to improve pitch performance through an in-tune aural guide and high speed visual feedback (Chapter 9): despite being dismissed as a useful mode for feedback, [136], [76, p.44,53], we found evidence that providing an in-tune aural guide was helpful for improving intonation based on both user reaction, and demonstrated effect on played pitch (Section 9.7). This suggests when designed with necessary consideration for how a musician listens and learns, aural feedback can be an appropriate mode for real-time feedback. Pedagogically relevant, though less popular, some beginners reported finding visual feedback more helpful, while students thought a combination would be the most helpful during practice.

10.3 Reflections on Complexity Management

One of our starting motivations in this thesis was looking at DMIs and recognizing the rarity of success attributable to some degree by the constant battle between initial playability and potential for virtuosity. Reviewing what is necessary in instrument design, Sergi Jorda comments [77]:

> "new digital instruments design is a quite broad subject, which includes highly technological areas (e.g. electronics and sensor technology, sound synthesis and processing techniques, computer programming...), human related disciplines (associated with psychology, ergonomics and many human-computer interaction components), plus all the possible connections between them (e.g. mapping techniques...). Low-level and focused research that tries to solve independent parts of the problem is clearly essential for any real progression in this field, but it is also clearly insufficient. Integral studies and approaches, which consider not only ergonomic but also psychological, philosophical and obviously, musical issues, even if non-systematic by definition, are also needed; but the fact is that very few attempts are being made at studying the design of new musical instruments -tools for playing and making music- as a conceptual whole."

While this is obviously a broad commentary, Jorda fails to include deep consideration of how a person learns. There is no mention of pedagogy, andragogy (study of teaching adults), or heutagogy (study of self-learning). Such oversight is common yet how a person learns an instrument, through chunking, practice and embedding, is how musicians became musicians and what any potential performer must progress through. Traditional learning succeeds in generating virtuosi. As Dobrian states: "The lack of virtuosity on new musical interfaces is apparently another case of the 'elephant in the corner,' a big bothersome issue that apparently everyone knows is there but is hesitant to discuss" [39].

Without focus on learning, few otherwise rich instruments will be successful. Some of the failure to include consideration for learning into instrument design is that achievement through practice is a long-term process. Timelines for academic research do not encourage study of long-term progress. Research like Cannon and Favilla's on the Bent Leather Band [23] are rare and often did not start as academic projects. Practice requires time commitment by users and testers, something that is a premium both within the research environment and outside of it.

There is also a certain element of hope for an ideal instrument that is virtuosic and expressive but does not require practice. More realistically, Dobrian [39] suggests that:

> for an instrument to be considered potentially expressive by a trained musician, it must necessarily have a certain degree of complexity in the relationship between input control data and sonic result. So it is reasonable to expect that such an instrument will have a certain learning curve; a performer will require a certain amount of training and practice to achieve good control of it. For high-quality musical expression, an instrument should be mastered; the performer should achieve a level of virtuosity.

Rather than seeing the balance between initial playability and potential for virtuosity as an either/or situation, we saw it as a conflict that can be resolved through expectations of practice and management of exposed complexity. We were inspired by game theory and how, similar to a new instrument, a game that is too easy may be fun at first but overall is not challenging and has limited longevity. Similarly the computer game that is too hard will fail to provide sufficient reward to sustain engagement no matter how rich it is otherwise. In games they solve the problem by intentionally designing the game to slowly increase in difficulty requiring the player to slowly develop additional skills to proceed in the game. Traditional instrument learning attempts to manage difficulty but is constrained by the inherent requirements of the instrument. DMIs still need to be learned, but an expressive, eventually virtuosic instrument need not suffer from overly complicated initial interactions. Rather, interactions can be designed to support different inherent levels of difficulty with each level offering more expressive capability and all levels offering appropriate musical reward.

10.4 Future Work

The augmented violin has proven itself both sufficiently useful and sufficiently similar to the violin for it to be useful in many further contexts, including performance and teaching. We are also interested in using it to further study the original idea of complexity management, which we will also discuss in a wider context beyond violins here.

10.4.1 The Augmented Violin

As with any hardware research design, there are always improvements that can made. Though the current fingerboard sensor design has proven robust and functional, there are opportunities to improve the build, especially adding shielding and investigating whether it can be made more sensitive to contact so that it tolerates lower finger pressure. The current finger pressure required is sufficient for participants with proper technique, but less tolerant of poor technique and vibrato, where pressure is often somewhat reduced when rocking the finger. If the sensor can be made robust in the presence of vibrato, that would be a useful improvement.

More significant improvements are needed for bow sensor mounting, and cabling could benefit from improved manufacturing. When it became apparent during use that physical sensor stability was key for maintaining tracking accuracy without requiring recalibration, we responded in an ad hoc manner that would not be suitable on a professional quality bow. Stable mounting is still a problem area. We will have more work to do before being able to augment a student's bow and send it home with them.

Additionally, if we want to send instruments home with students for practice, the system will need to be made physically smaller. Moving the aural and visual feedback systems onto a single platform like a BeagleBone[1] with Bela[2] [113] would ensure not only the lowest latency possible, but would make the augmented violin more portable. Appropriate packaging could make it more attractive and better secure cables. Additionally, full UI functionality is not necessary and software could be redesigned so that all pedagogical applications in this thesis only require a brief calibration stage and selection of key before students can use it. There are no inherent reasons why the augmented violin could not be improved so it can be sent home for longitudinal testing in student practice.

10.4.2 Intonation Learning Aids

Although we tested our intonation feedback methods in a lesson environment for practical reasons, our target environment is home practice. In the lesson context, the author/teacher was often tasking students to demonstrate immediate accomplishment or improvement on a specific task that was not necessarily pitch related. As a result, students may have focused on the most expedient and familiar way to accomplish the requested task which may have involved ignoring the provided feedback. Additionally, a repeated comment of study participants was that it took time to get comfortable with how

[1] http://beagleboard.org/bone
[2] http://www.bela.io/

to use the aural feedback. It also takes time to make a noticeable repeatable improvement in intonation. A single 30 minute lesson with a given feedback type is insufficient for significant improvement in pitch accuracy.

Home practice is where students could most benefit from technology to help identify error as well as how to correct it. A practice-based study would allow students the time to investigate and explore the feedback methods in a non-pressurized environment and also enable the repeated use that leads to learning. A longitudinal study of use during practice is necessary to better validate whether using our pitch feedback methods successfully improve intonation learning and are helpful enough for students to choose to use them regularly.

One of the more interesting findings of the study with beginners (Chapter 9) was that modes that highlighted error more strongly, either through visual illustration or conspicuous audio artifacts, caused a stronger desire in participants to fix the error. As one participant stated, though the aural feedback did provide the opportunity to listen for error, "because I can hear it right, it almost means I don't correct." This is exactly the reaction we do not want. It would be interesting to investigate whether intentionally exaggerating intonation error would yield a more beneficial practice tool.

A further effect that can only be followed up through a longitudinal study is whether learning that occurs using our feedback tools is sustained. In [136] having (incorrectly in our opinion) dismissed aural feedback as a viable option for learning feedback due to overlap with the primary instrument sound, Percival argues against providing real-time visual feedback to students as it becomes an artificial crutch; students learn to rely on feedback poorly related to normal performance tasks. In order to assess the true usefulness of both aural and visual feedback it is necessary to study whether the positive effects of using either learning tool continue when the tool is removed.

10.4.3 Simplifying the Violin

As we stated in Chapter 2, attempting to prove the usefulness of complexity management by designing a new virtuosic instrument that we could optimize for different difficulties and then testing how the design impacted long term practice, learning, and practice motivation was overly risky; it is too easy to design what turns out to be a bad instrument. For this reason we chose to begin investigations by simplifying the violin, however complexity management is an idea with implications that extend beyond the violin.

In this thesis we completed the major precursors needed for performing a more significant longitudinal study on complexity management using the violin. Pitch was chosen as the target for intervention not only because it was a major learning task but it was easier to understand how to control improvements to heard pitch through pitch correction. In our study with beginners (Chapter 9) we verified a major assumption of complexity management, that beginners would responded positively to and enjoy a simplified pitch corrected experience. We also found that although experts disliked pitch correction due to it interrupting their internalized intonation process loop (Section 2.4.2), allowing limited control by partially retuning the audio without fully snapping to the nearest semitone resulted in an improved experience not significantly worse than when there is no correction. Although much work remains to design and implement the necessary studies to fully evaluate the usefulness, effects and pacing of managing complexity in pitch, both results strongly suggest that pitch simplification is a suitable target for research. Pitch snapping could be like a training wheel that is slowly removed as the student demonstrates improved intonation.

A fortuitous discovery during the study with beginners was that heavy dynamic range compression of violin audio reduced the chance for the audibility of bow error to propagate to heard sound. Controlling compression level effectively provides us with means for managing bow complexity. We are curious not just about pitch simplification but also bow simplification and the two together. For instance, as we stated initially, part of the frustration when learning violin is that focusing on bow and pitch tasks separately can only be done effectively during extremely simplified and dull tasks. Playing repertoire, there are too many left hand demands to effectively focus solely on bow technique. However, if we simplify pitch, will that allow students to focus more directly on bowing and vice versa? With these two ways of simplifying violin performance, an inexpensive real-time augmented violin, and our preliminary findings, we are in a strong position for continuing research into complexity management using the violin and conducting studies with a longer, more realistic timeline for monitoring student progression and intervention impact.

Bibliography

[1] Bernard D Adelstein, Durand R Begault, Mark R Anderson, and Elizabeth M Wenzel. Sensitivity to haptic-audio asynchrony. In *Proceedings of the 5th international conference on Multimodal interfaces*, pages 73–76. ACM, 2003.

[2] A. Askenfelt. Measurement of bow motion and bow force in violin playing. *Journal Acoustical Society of America*, 80:1007, 1986.

[3] A. Askenfelt. Measurement of the bowing parameters in violin playing. ii: Bow–bridge distance, dynamic range, and limits of bow force. *Journal Acoustical Society of America*, 86:503, 1989.

[4] A. Askenfelt. Observations on the violin bow and the interaction with the string. In *Proceedings of the 1995 International Symposium on Musical Acoustics*, 1995.

[5] Curtis Bahn and Dan Trueman. Interface: electronic chamber ensemble. In *Proceedings of the 2001 Conference on New Interfaces for Musical Expression*, pages 1–5, 2001.

[6] Emmanouil Benetos. *Automatic transcription of polyphonic music exploiting temporal evolution.* PhD thesis, Queen Mary University of London, 2012.

[7] Frédéric Bevilacqua, Jeff Ridenour, and David J Cuccia. 3d motion capture data: motion analysis and mapping to music. In *Proceedings of the Workshop/symposium on Sensing and Input for Media-centric Systems*, 2002.

[8] Frédéric Bevilacqua, Nicolas Rasamimanana, Emmanuel Fléty, Serge Lemouton, and Florence Baschet. The augmented violin project: research, composition and performance report. In *Proceedings of the 2006 Conference on New Interfaces for Musical Expression*, pages 402–406, 2006.

[9] Christopher M Bishop and Nasser M Nasrabadi. *Pattern recognition and machine learning*, volume 1. springer New York, 2006.

[10] Renaud Blanch, Yves Guiard, and Michel Beaudouin-Lafon. Semantic pointing: improving target acquisition with control-display ratio adaptation. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 519–526. ACM, 2004.

[11] Sebastian Böck and Gerhard Widmer. Maximum filter vibrato suppression for onset detection. In *Proceedings of the 16th International Conference on Digital Audio Effects (DAFX-13)*, 2013.

[12] Sebastian Böck, Andreas Arzt, Florian Krebs, and Markus Schedl. Online realtime onset detection with recurrent neural networks. In *Proceedings of the 15th International Conference on Digital Audio Effects (DAFX-12)*, page 4, 2012.

[13] Sebastian Böck, Florian Krebs, and Markus Schedl. Evaluating the online capabilities of onset detection methods. In *Proceedings of the 13th Conference of the International Society for Music Information Retrieval (ISMIR)*, pages 49–54, 2012.

[14] Bert Bongers. Physical interfaces in the electronic arts. *Trends in Gestural Control of Music*, pages 41–70, 2000.

[15] Adam Bowen. Soundstone: a 3-d wireless music controller. In *Proceedings of the 2005 Conference on New Interfaces for Musical Expression*, pages 268–269, 2005.

[16] Virginia Braun and Victoria Clarke. Using thematic analysis in psychology. *Qualitative research in psychology*, 3(2):77–101, 2006.

[17] Judith C Brown. Musical fundamental frequency tracking using a pattern recognition method. *Journal Acoustical Society of America*, 92:1394, 1992.

[18] Judith C Brown and Kathryn V Vaughn. Pitch center of stringed instrument vibrato tones. *The Journal of the Acoustical Society of America*, 100(3):1728–1735, 1996.

[19] Anne-Marie Burns and Marcelo M Wanderley. Visual methods for the retrieval of guitarist fingering. In *Proceedings of the 2006 Conference on New Interfaces for Musical Expression*, pages 196–199, 2006.

[20] William Buxton and Roger Dannenberg. The computer as musical accompanist. *ACM SIGCHI Bulletin*, 17(4):41–43, 1986.

[21] Ozan Cakmakci, François Bérard, and Joëlle Coutaz. An augmented reality based learning assistant for electric bass guitar. In *Proc. of the 10th International Conference on Human-Computer Interaction, Crete, Greece*, 2003.

[22] Chris Cannam, Christian Landone, Mark B Sandler, and Juan Pablo Bello. The sonic visualiser:

A visualisation platform for semantic descriptors from musical signals. In *Proceedings of the 7th Conference of the International Society for Music Information Retrieval (ISMIR)*, pages 324–327, 2006.

[23] Joanne Cannon and Stuart Favilla. *The investment of play: expression and affordances in digital musical instrument design*. Ann Arbor, MI: Michigan Publishing, University of Michigan Library, 2012.

[24] Alfonso Perez Carrillo. Characterization of bowing strokes in violin playing in terms of controls and sound: Differences between bouncing and on-string bow strokes. In *Proceedings of Meetings on Acoustics*, volume 19. Acoustical Society of America, 2013.

[25] Jessie Chen, Marjorie H Woollacott, Steven Pologe, and George P Moore. Pitch and space maps of skilled cellists: accuracy, variability, and error correction. *Experimental brain research*, 188 (4):493–503, 2008.

[26] Magdalena Chudy, Alfonso Perez Carrillo, and Simon Dixon. On the relation between gesture, tone production and perception in classical cello performance. In *Proceedings of Meetings on Acoustics*, volume 19. Acoustical Society of America, 2013.

[27] Nick Collins. A comparison of sound onset detection algorithms with emphasis on psychoacoustically motivated detection functions. In *Audio Engineering Society Convention 118*. Audio Engineering Society, 2005.

[28] Nick Collins. Using a pitch detector for onset detection. In *Proceedings of the 6th Conference of the International Society for Music Information Retrieval (ISMIR)*, pages 100–106, 2005.

[29] Arshia Cont. Antescofo: Anticipatory synchronization and control of interactive parameters in computer music. In *Proceedings of the 2008 International Computer Music Conference*, 2008.

[30] Perry Cook. Principles for designing computer music controllers. In *Proceedings of the 2001 Conference on New Interfaces for Musical Expression*, pages 1–4, 2001.

[31] Perry Cook. Re-designing principles for computer music controllers: A case study of squeezevox maggie. In *Proceedings of the 2009 Conference on New Interfaces for Musical Expression*, pages 218–221, 2009.

[32] Mihaly Csikszentmihalyi. *Flow and the psychology of discovery and invention*. Harper Collins, 1996.

[33] Alain de Cheveigné and Hideki Kawahara. Comparative evaluation of f0 estimation algorithms. In *INTERSPEECH*, pages 2451–2454, 2001.

[34] Alain De Cheveigné and Hideki Kawahara. Yin, a fundamental frequency estimator for speech and music. *Journal Acoustical Society of America*, 111:1917, 2002.

[35] François De Sorbier, Hiroyuki Shiino, and Hideo Saito. Violin pedagogy for finger and bow placement using augmented reality. In *Signal & Information Processing Association Annual Summit and Conference (APSIPA ASC), 2012 Asia-Pacific*, pages 1–5. IEEE, 2012.

[36] Matthias Demoucron. *On the control of virtual violins-Physical modelling and control of bowed string instruments*. PhD thesis, Université Pierre et Marie Curie-Paris VI, 2008.

[37] Matthias Demoucron, Anders Askenfelt, and René Caussé. Measuring bow force in bowed string performance: Theory and implementation of a bow force sensor. *Acta Acustica united with Acustica*, 95(4):718–732, 2009.

[38] Simon Dixon. Onset detection revisited. In *Proc. DAFx*, pages 133–137, 2006.

[39] Christopher Dobrian and Daniel Koppelman. The'e'in nime: musical expression with new computer interfaces. In *Proceedings of the 2006 Conference on New Interfaces for Musical Expression*, pages 277–282, 2006.

[40] Ross W Duffin. *How equal temperament ruined harmony (and why you should care)*. WW Norton & Company, 2008.

[41] K Anders Ericsson, Ralf T Krampe, and Clemens Tesch-Römer. The role of deliberate practice in the acquisition of expert performance. *Psychological review*, 100(3):363, 1993.

[42] Rebecca Fiebrink, Dan Trueman, and Perry R Cook. A metainstrument for interactive, on-the-fly machine learning. In *Proceedings of the 2009 Conference on New Interfaces for Musical Expression*, volume 2, page 3, 2009.

[43] Stevena Finney and Caroline Palmer. Auditory feedback and memory for music performance: Sound evidence for an encoding effect. *Memory & cognition*, 31(1):51–64, 2003.

[44] Lawrence M. Fisher. Ivories that tickle themselves. *The New York Times*, February 1993.

[45] Neil Fleming and David Baume. Learning styles again: Varking up the right tree! *Educational Developments*, 7(4):4, 2006.

[46] D Fober, S Letz, and Y Orlarey. Vemus-feedback and groupware technologies for music instrument learning. In *Proceedings of the 4th Sound and Music Computing Conference (SMC)*, volume 7, pages 117–123, 2007.

[47] Ivan Franco. The airstick: a free-gesture controller using infrared sensing. In *Proceedings of the 2005 Conference on New Interfaces for Musical Expression*, pages 248–249, 2005.

[48] A. Freed. Novel and forgotten current-steering techniques for resistive multitouch, duotouch, and polytouch position sensing with pressure. In *Proceedings of the 2009 Conference on New Interfaces for Musical Expression*, 2009.

[49] Adrian Freed, Amar Chaudhary, and Brian Davila. Operating systems latency measurement and analysis for sound synthesis and processing applications. In *Proceedings of the 1997 International Computer Music Conference*, pages 479–81, 1997.

[50] Adrian Freed, David Wessel, Michael Zbyszynski, and Frances Marie Uitti. Augmenting the cello. In *Proceedings of the 2006 Conference on New Interfaces for Musical Expression*, pages 409–413, 2006.

[51] Adrian Freed, Frances-Marie Uitti, Sam Mansfield, and John MacCallum. "Old" is the new "New": a fingerboard case study in recrudescence as a NIME development strategy. In *Proceedings of the 2013 Conference on New Interfaces for Musical Expression*, 2013.

[52] Ivan Galamian. *Principles of violin playing and teaching.* Prentice Hall, Inc., 1962.

[53] John Geake. Neuromythologies in education. *Educational Research*, 50(2):123–133, 2008.

[54] John M Geringer, Rebecca B MacLeod, and Julia C Ellis. Two studies of pitch in string instrument vibrato: Perception and pitch matching responses of university and high school string players. *International Journal of Music Education*, 32(1):19–30, 2014.

[55] Dan Gibson. The modified cello. In *Proceedings of 2016 Conference on International Conference on Live Interfaces (ICLI)*, 2016.

[56] Marco Gillies, Harry Brenton, Matthew Yee-King, Andreu Grimalt-Reynes, and Mark d'Inverno. Sketches vs skeletons: video annotation can capture what motion capture cannot. In *Proceedings of the 2nd International Workshop on Movement and Computing*, pages 104–111. ACM, 2015.

[57] Suguru Goto. The aesthetics and technological aspects of virtual musical instruments: The case of the superpolm midi violin. *Leonardo Music Journal*, 9:115–120, 1999.

[58] Tobias Grosshauser and Thomas Hermann. Augmented haptics–an interactive feedback system for musicians. In *International Conference on Haptic and Audio Interaction Design*, pages 100–108. Springer, 2009.

[59] Tobias Grosshauser and Thomas Hermann. The sonified music stand–an interactive sonification

system for musicians. In *Proceedings of the 6th Sound and Music Computing Conference (SMC)*, pages 233–238. Casa da Musica, Porto, Portugal, 2009.

[60] Tobias Grosshauser and Gerhard Tröster. Further finger position and pressure sensing techniques for stringed and keyboard instruments. In *Proceedings of the 2013 Conference on New Interfaces for Musical Expression*, 2013.

[61] Tobias Grosshauser, V Candia, H Hildebrandt, and G Tröster. Sensor based measurements of musicians' synchronization issues. In *Proceedings of the 2012 Conference on New Interfaces for Musical Expression*, 2012.

[62] Tobias Grosshauser, Sebastian Feese, and Gerhard Tröster. Capacitive left hand finger and bow sensors for synchronization and rhythmical regularity analysis in string ensembles. In *Proceedings of the 2013 Stockholm Music Acoustics Conference, SMAC*, 2013.

[63] Enric Guaus, Jordi Bonada, Alfonso Perez, Esteban Maestre, and Merlijn Blaauw. Measuring the bow pressing force in a real violin performance. In *Proceedings of International Symposium on Musical Acoustics*, 2007.

[64] Enric Guaus, Jordi Bonada, Esteban Maestre, Alfonso Pérez, and Merlijn Blaauw. Calibration method to measure accurate bow force for real violin performances. In *Proceedings of the 2009 International Computer Music Conference*, 2009.

[65] James VM Hanson, David Whitaker, and James Heron. Preferential processing of tactile events under conditions of divided attention: Effects of divided attention on reaction time. *Neuroreport*, 20(15):1392, 2009.

[66] Jordan Hochenbaum and Ajay Kapur. Toward the future practice room: Empowering musical pedagogy through hyperinstruments. In *Proceedings of the 2013 Conference on New Interfaces for Musical Expression*, pages 307–312, 2013.

[67] Simon Holland, Anders J Bouwer, Mathew Dalgelish, and Topi M Hurtig. Feeling the beat where it counts: fostering multi-limb rhythm skills with the haptic drum kit. In *Proceedings of the 4th International Conference on Tangible, Embedded, and Embodied Interaction (TEI '10)*, pages 21–28. ACM, 2010.

[68] Lu Huanhuan. *Computer assisted music instrument tutoring applied to violin practice.* PhD thesis, National University of Singapore, 2009.

[69] Andy Hunt and Ross Kirk. Mapping strategies for musical performance. *Trends in Gestural Control of Music*, 21, 2000.

[70] Andy Hunt, Marcelo Wanderley, and Ross Kirk. Towards a model for instrumental mapping in expert musical interaction. In *Proceedings of the 2000 International Computer Music Conference*, pages 209–212, 2000.

[71] Andy Hunt, Marcelo M Wanderley, and Matthew Paradis. The importance of parameter mapping in electronic instrument design. *Journal New Music Research*, 32(4):429–440, 2003.

[72] Robert Jack, Andrew McPherson, and Tony Stockman. Effect of latency on performer interaction and subjective quality assessment of a digital musical instrument. In *AM '16: Proceedings of the 10th Audio Mostly: A Conference on Interaction With Sound*. ACM, 2016.

[73] R Jensenius and V. Johnson. Performing the electric violin in a sonic space. *Computer Music Journal*, 36(4):28–39, 2012.

[74] Rose Johnson, Yvonne Rogers, Janet van der Linden, and Nadia Bianchi-Berthouze. Being in the thick of in-the-wild studies: the challenges and insights of researcher participation. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 1135–1144. ACM, 2012.

[75] Rose Johnson, Nadia Bianchi-Berthouze, Yvonne Rogers, and Janet van der Linden. Embracing calibration in body sensing: using self-tweaking to enhance ownership and performance. In *Proceedings of the 2013 ACM International Joint Conference on Pervasive and Ubiquitous Computing*, pages 811–820. ACM, 2013.

[76] Rose Mary Grace Johnson. *In Touch with the Wild: Exploring Real-time Feedback for Learning to Play the Violin*. PhD thesis, UCL (University College London), 2014.

[77] Sergi Jordà. Digital instruments and players: part i—efficiency and apprenticeship. In *Proceedings of the 2004 Conference on New Interfaces for Musical Expression*, pages 59–63, 2004.

[78] Sergi Jordà. Instruments and players: Some thoughts on digital lutherie. *Journal New Music Research*, 33(3):321–341, 2004.

[79] Pirkko Juntunen. Music technology promoting violin and string instrument instruction. *CFMAE Interdisciplinary Journal of Music and Art Pedagogy*, 3:17–34, 2011.

[80] Pirkko Juntunen, H Ruismäki, and I Ruokonen. Music technology in finnish string instrument and orchestra instruction. In *Design Learning and Well-being: 4th International Journal of Intercultural Arts Education: Post-Conference Book: University of Helsinki, Department of Teacher Education. Research report*, volume 331, 2011.

[81] Martin Kaltenbrunner, Sergi Jorda, Gunter Geiger, and Marcos Alonso. The reactable*: A collaborative musical instrument. In *Enabling Technologies: Infrastructure for Collaborative Enterprises, 2006. WETICE'06. 15th IEEE International Workshops on*, pages 406–411. IEEE, 2006.

[82] Ajay Kapur, Ariel J Lazier, Philip Davidson, R Scott Wilson, and Perry R Cook. The electronic sitar controller. In *Proceedings of the 2004 Conference on New Interfaces for Musical Expression*, 2004.

[83] Ajay Kapur, George Tzanetakis, Naznin Virji-Babul, Ge Wang, and Perry R Cook. A framework for sonification of vicon motion capture data. In *Proceedings of the 8th International Conference on Digital Audio Effects (DAFX-05)*, pages 47–52, 2005.

[84] Mark Katz. *Capturing sound: how technology has changed music*. Univ of California Press, 2010.

[85] Susan Kempter. *How muscles learn: teaching the violin with the body in mind*. Alfred Music Publishing, 2003.

[86] Chris Kiefer, Nick Collins, and Geraldine Fitzpatrick. Hci methodology for evaluating musical controllers: A case study. In *Proceedings of 2008 Conference on New Interfaces for Musical Expression*, volume 8, 2008.

[87] Anssi Klapuri. Sound onset detection by applying psychoacoustic knowledge. In *Acoustics, Speech, and Signal Processing, 1999. Proceedings., 1999 IEEE International Conference on*, volume 6, pages 3089–3092. IEEE, 1999.

[88] Daniel L Kohut. *Musical performance: Learning theory and pedagogy*. Prentice-Hall New Jersey, 1985.

[89] Jürgen Konczak, Heidi vander Velden, and Lukas Jaeger. Learning to play the violin: motor control by freezing, not freeing degrees of freedom. *Journal of motor behavior*, 41(3):243–252, 2009.

[90] Nina Kraus and Bharath Chandrasekaran. Music training for the development of auditory skills. *Nature Reviews Neuroscience*, 11(8):599–605, 2010.

[91] Edward Kreitman. *Teaching from the Balance Point: A Guide for Suzuki Parents, Teachers, and Students*. Western Springs School of Talent Education, 1998.

[92] Edward Kreitman. Teaching with an open heart: A guide to developing conscious musicianship for suzuki parents, teachers, and students, 2010.

[93] M Lahat, R Niederjohn, and D Krubsack. A spectral autocorrelation method for measurement of the fundamental frequency of noise-corrupted speech. *Acoustics, Speech and Signal Processing, IEEE Transactions on*, 35(6):741–750, 1987.

[94] Oliver Larkin, Thijs Koerselman, Bee Ong, and Kia Ng. Sonification of bowing features for string instrument training. In *Proceedings of the 14th International Conference on Auditory Display*, 2008.

[95] N. Leroy, E. Fléty, and F. Bevilacqua. Reflective optical pickup for violin. In *Proceedings of the 2006 Conference on New Interfaces for Musical Expression*, pages 204–207, 2006.

[96] Michael Lester and Jon Boley. The effects of latency on live sound monitoring. In *Audio Engineering Society Convention 123*. Audio Engineering Society, 2007.

[97] Daniel J Levitin, Stephen McAdams, and Robert L Adams. Control parameters for musical instruments: a foundation for new mappings of gesture to sound. *Organised Sound*, 7(2):171–189, 2002.

[98] Kyung Ae Lim and Christopher Raphael. Intune: A system to support an instrumentalist's visualization of intonation. *Computer Music Journal*, 34(3):45–55, 2010.

[99] Markus Löchtefeld, Sven Gehring, Ralf Jung, and Antonio Krüger. Using mobile projection to support guitar learning. In *International Symposium on Smart Graphics*, pages 103–114. Springer, 2011.

[100] Huanhuan Lu, Bingjun Zhang, Ye Wang, and Wee Kheng Leow. idvt: An interactive digital violin tutoring system based on audio-visual fusion. In *Proceedings of the 16th ACM international Conference on Multimedia*, pages 1005–1006. ACM, 2008.

[101] E. Maestre, J. Bonada, M. Blaauw, A. Perez, and E. Guaus. Acquisition of violin instrumental gestures using a commercial EMF tracking device. In *Proceedings of the 2007 International Computer Music Conference*, volume 1, pages 386–393, 2007.

[102] Thor Magnusson. Designing constraints: Composing and performing with digital musical systems. *Computer Music Journal*, 34(4):62–73, 2010.

[103] Thor Magnusson and Enrike Hurtado Mendieta. The acoustic, the digital and the body: A survey on musical instruments. In *Proceedings of the 2007 Conference on New Interfaces for Musical Expression*, pages 94–99. ACM, 2007.

[104] Clemens Maidhof. Error monitoring in musicians. *Frontiers in human neuroscience*, 7:401, 2013.

[105] Teemu Mäki-Patola. Musical effects of latency. *Suomen Musiikintutkijoiden*, 9:82–85, 2005.

[106] Teemu Mäki-Patola and Perttu Hämäläinen. Latency tolerance for gesture controlled continuous sound instrument without tactile feedback. In *ICMC*, 2004.

[107] Joseph Malloch and Marcelo M Wanderley. The t-stick: from musical interface to musical instrument. In *Proceedings of the 2007 Conference on New Interfaces for Musical Expression*, pages 66–70. ACM, 2007.

[108] Marco Marchini, Panagiotis Papiotis, Alfonso Pérez, and Esteban Maestre. A hair ribbon deflection model for low-intrusiveness measurement of bow force in violin performance. In *Proceedings of the 2011 Conference on New Interfaces for Musical Expression*, pages 481–486, 2011.

[109] Max V Mathews and J Kohut. Electronic simulation of violin resonances. *Journal Acoustical Society of America*, 53:1620, 1973.

[110] Keith A McMillen. Stage-worthy sensor bows for stringed instruments. In *Proceedings of the 2008 Conference on New Interfaces for Musical Expression*, pages 347–348, 2008.

[111] A. McPherson and Y. Kim. Augmenting the acoustic piano with electromagnetic string actuation and continuous key position sensing. In *Proceedings of the 2010 Conference on New Interfaces for Musical Expression*, 2010.

[112] Andrew McPherson. Portable measurement and mapping of continuous piano gesture. In *Proceedings of the 2010 Conference on New Interfaces for Musical Expression*, pages 152–157, 2013.

[113] Andrew McPherson and Victor Zappi. An environment for submillisecond-latency audio and sensor processing on beaglebone black. In *Audio Engineering Society Convention 138*. Audio Engineering Society, 2015.

[114] Duncan Menzies. *Technological Support for Highland Piping Tuition and Practice*. PhD thesis, Queen Mary University of London, 2015.

[115] Dylan Menzies, Brian Bramer, Dilip Chauhan, and Mark Cuckow. New developments in the o-bow. In *Proceedings of the fourth international Conference on Tangible, Embedded, and Embodied Interaction (TEI '13)*. ACM, 2013.

[116] David Merrill and Joseph A Paradiso. Personalization, expressivity, and learnability of an implicit mapping strategy for physical interfaces. In *Proceedings of the CHI Conference on Human Factors in Computing Systems, Extended Abstracts*, pages 2152–2161, 2005.

[117] Christophe Micheyl, Karine Delhommeau, Xavier Perrot, and Andrew J Oxenham. Influence of

musical and psychoacoustical training on pitch discrimination. *Hearing research*, 219(1):36–47, 2006.

[118] George A Miller. Some psychological studies of grammar. *American psychologist*, 17(11):748, 1962.

[119] Yoichi Motokawa and Hideo Saito. Support system for guitar playing using augmented reality display. In *2006 IEEE/ACM International Symposium on Mixed and Augmented Reality*, pages 243–244. IEEE, 2006.

[120] Kia Ng and Paolo Nesi. i-maestro: Technology-enhanced learning and teaching for music. In *Proceedings of the 2008 Conference on New Interfaces for Musical Expression*, pages 225–228, 2008.

[121] Kia Ng, Oliver Larkin, Thijs Koerselman, and Bee Ong. 3d motion data analysis and visualisation for string practice training. In *Electronic Visualisation and the Arts (EVA) London Conference*, volume 197, 2007.

[122] Kia C Ng, Tillman Weyde, Oliver Larkin, Kerstin Neubarth, Thijs Koerselman, and Bee Ong. 3d augmented mirror: a multimodal interface for string instrument learning and teaching with gesture support. In *Proceedings of the 9th international Conference on Multimodal interfaces*, pages 339–345. ACM, 2007.

[123] A Michael Noll. Cepstrum pitch determination. *Journal Acoustical Society of America*, 41:293, 1967.

[124] Joseph O'Connor. *Not pulling strings*. Metamorphous Press, 1987.

[125] Dan Overholt. The Overtone Violin: A new computer music instrument. In *Proceedings of the 2005 International Computer Music Conference*, pages 604–607, 2005.

[126] Dan Overholt. Violin-related HCI: A taxonomy elicited by the musical interface technology design space. In *Arts and Technology*, pages 80–89. Springer, 2012.

[127] Dan Overholt. Advancements in violin-related human-computer interaction. *International Journal of Arts and Technology 2*, 7(2-3):185–206, 2014.

[128] Daniel Overholt. The Overtone Fiddle: an actuated acoustic instrument. In *Proceedings of the 2011 Conference on New Interfaces for Musical Expression*, 2011.

[129] J.A. Paradiso and N. Gershenfeld. Musical applications of electric field sensing. *Computer Music Journal*, 21(2):69–89, 1997.

[130] Joseph A Paradiso. Electronic music: new ways to play. *Spectrum, IEEE*, 34(12):18–30, 1997.

[131] Laurel S Pardue and Andrew P McPherson. Near-field optical reflective sensing for bow tracking. In *Proceedings of the 2013 Conference on New Interfaces for Musical Expression*, 2013.

[132] Laurel S Pardue and Joseph A Paradiso. Musical navigatrics: new musical interactions with passive magnetic tags. In *Proceedings of the 2002 Conference on New Interfaces for Musical Expression*, pages 1–3, 2002.

[133] Laurel S Pardue, Dongjuan Nian, Chris Harte, and Andrew P McPherson. Low-latency audio pitch tracking: a multi-modal sensor-assisted approach. In *Proceedings of the 2014 Conference on New Interfaces for Musical Expression*, 2014.

[134] Laurel S Pardue, Christopher Harte, and Andrew P McPherson. A low-cost real-time tracking system for violin. *Journal of New Music Research*, 44(4):305–323, 2015.

[135] Chad Peiper, David Warden, and Guy Garnett. An interface for real-time classification of articulations produced by violin bowing. In *Proceedings of the 2003 Conference on New Interfaces for Musical Expression*, pages 192–196, 2003.

[136] Graham Percival, Ye Wang, and George Tzanetakis. Effective use of multimedia for computer-assisted musical instrument tutoring. In *Proceedings of the international workshop on Educational multimedia and multimedia education*, pages 67–76. ACM, 2007.

[137] Olivier Perrotin and Christophe d'Alessandro. Adaptive mapping for improved pitch accuracy on touch user interfaces. In *Proceedings of the 2013 Conference on New Interfaces for Musical Expression*, pages 186–189, 2013.

[138] Peter Q Pfordresher and Caroline Palmer. Effects of hearing the past, present, or future during music performance. *Attention, Perception, & Psychophysics*, 68(3):362–376, 2006.

[139] Cornelius Poepel. On interface expressivity: a player-based study. In *Proceedings of the 2005 Conference on New Interfaces for Musical Expression*, pages 228–231, 2005.

[140] Cornelius Poepel and Dan Overholt. Recent developments in violin-related digital musical instruments: where are we and where are we going? In *Proceedings of the 2006 Conference on New Interfaces for Musical Expression*, 2006.

[141] Lawrence Rabiner. On the use of autocorrelation analysis for pitch detection. *Acoustics, Speech and Signal Processing, IEEE Transactions on*, 25(1):24–33, 1977.

[142] R. Ramirez, A. Perez, S. Kersten, and E. Maestre. Performer identification in celtic violin

recordings. In *Proc. of 9th International Conference on Music Information Retrieval*, pages 483–488, 2008.

[143] Spyros Raptis, Aimilios Chalamandaris, Alexandros Baxevanis, A Askenfeld, Erwin Schoonderwaldt, K Falkenberg Hansen, Dominique Fober, Stéphane Letz, and Yann Orlarey. Imutus-an effective practicing environment for music tuition. In *Proceedings of the 2005 International Computer Music Conference*, pages 383–386, 2005.

[144] Nicolas Rasamimanana, Emmanuel Fléty, and Frédéric Bevilacqua. Gesture analysis of violin bow strokes. *Gesture in Human-Computer Interaction and Simulation*, pages 145–155, 2006.

[145] James Reel. 14 steps towards improved intonation. *Strings*, Oct 2014. http://stringsmagazine.com/14-steps-toward-improved-intonation/.

[146] Bruno H Repp. Effects of auditory feedback deprivation on expressive piano performance. *Music Perception: An Interdisciplinary Journal*, 16(4):409–438, 1999.

[147] Roland D-Beam. http://www.roland.com/products/en/exp/D_BEAM.html.

[148] Edward Rothstein. Yo-Yo Ma and his new 'Hyper' Cello. *The New York Times*, August 1991.

[149] Joseph Butch Rovan, Marcelo M Wanderley, Shlomo Dubnov, and Philippe Depalle. Instrumental gestural mapping strategies as expressivity determinants in computer music performance. In *Proceedings of Kansei-The Technology of Emotion Workshop*, pages 3–4, 1997.

[150] María Herrojo Ruiz, Hans-Christian Jabusch, and Eckart Altenmüller. Detecting wrong notes in advance: neuronal correlates of error monitoring in pianists. *Cerebral Cortex*, 19(11):2625–2639, 2009.

[151] E. Schoonderwaldt. *Mechanics and acoustics of violin bowing: Freedom, constraints and control in performance*. PhD thesis, KTH Royal Institute of Technology, 2009.

[152] E. Schoonderwaldt. The player and the bowed string: Coordination of bowing parameters in violin and viola performance. *Journal Acoustical Society of America*, 126:2709, 2009.

[153] E. Schoonderwaldt, K. Guettler, and A. Askenfelt. Effect of the width of the bow hair on the violin string spectrum. In *Proceedings of the 2003 Stockholm Music Acoustics Conference, SMAC*, 2003.

[154] Erwin Schoonderwaldt and Eckart Altenmüller. Mastering the violin: Motor learning in complex bowing skills. In *Proceedings of the International Symposium on Performance Science*, pages 649–654, 2011.

[155] Erwin Schoonderwaldt and Matthias Demoucron. Extraction of bowing parameters from violin performance combining motion capture and sensors. *Journal Acoustical Society of America*, 126: 2695, 2009.

[156] Erwin Schoonderwaldt and Alexander Refsum Jensenius. Effective and expressive movements in a French-Canadian fiddler's performance. In *Proceedings of the 2011 Conference on New Interfaces for Musical Expression*, 2011.

[157] Erwin Schoonderwaldt and M Wanderley. Visualization of bowing gestures for feedback: The hodgson plot. In *Proceedings of the 3rd International Conference on Automated Production of Cross Media Content for Multi-channel Distribution (AXMEDIS07)*, volume 2, pages 65–70, 2007.

[158] Erwin Schoonderwaldt, Kjetil Hansen, and Anders Askenfeld. Imutus–an interactive system for learning to play a musical instrument. In *Proceedings of the International Conference of Interactive Computer Aided Learning (ICL)*, pages 143–150, 2004.

[159] John G Sharp, Rob Bowker, and Jenny Byrne. Vak or vak-uous? towards the trivialisation of learning and the death of scholarship. *Research Papers in Education*, 23(3):293–314, 2008.

[160] John A Sloboda, Jane W Davidson, Michael JA Howe, and Derek G Moore. The role of practice in the development of performing musicians. *British Journal of Psychology*, 87(2):287–309, 1996.

[161] Yuji Sone et al. Sensory otherness in laurie anderson's work. *Body, Space, &Technology Journal*, 5, 2005.

[162] William Starr and Constance Starr. *To learn with love: A companion for Suzuki parents*. Alfred Music, 1983.

[163] Donovan Stokes. Challenges to good intonation in a live situation. *No Treble*, February 2015.

[164] Dan Stowell and Mark Plumbley. Adaptive whitening for improved real-time audio onset detection. In *Proceedings of the 2007 International Computer Music Conference*, volume 18, 2007.

[165] Shinichi Suzuki and Waltraud Suzuki. *Nurtured by love: The classic approach to talent education*. Alfred Music, 1983.

[166] Dan Trueman and Perry Cook. Bossa: The deconstructed violin reconstructed. *Journal of New Music Research*, 29(2):121–130, 2000.

[167] Josh Tyrangiel. Auto-tune: Why pop music sounds perfect. *Time Magazine*, pages 1877372–3, 2009.

[168] Janet Van Der Linden, Rose Johnson, Jon Bird, Yvonne Rogers, and Erwin Schoonderwaldt. Buzzing to play: lessons learned from an in the wild study of real-time vibrotactile feedback. In *Proceedings of the SIGCHI Conference on Human factors in Computing Systems*, pages 533–542. ACM, 2011.

[169] Janet van der Linden, Erwin Schoonderwaldt, Jon Bird, and Rose Johnson. Musicjacket: combining motion capture and vibrotactile feedback to teach violin bowing. *Instrumentation and Measurement, IEEE Transactions on*, 60(1):104–113, 2011.

[170] Werner Verhelst and Marc Roelands. An overlap-add technique based on waveform similarity (wsola) for high quality time-scale modification of speech. In *Acoustics, Speech, and Signal Processing, 1993. ICASSP-93., 1993 IEEE International Conference on*, volume 2, pages 554–557. IEEE, 1993.

[171] Adrian von dem Knesebeck and Udo Zölzer. Comparison of pitch trackers for real-time guitar effects. In *Proc. DAFx*, pages 266–269, 2010.

[172] Marcelo M Wanderley. Gestural control of music. In *International Workshop Human Supervision and Control in Engineering and Music*, pages 632–644, 2001.

[173] Marcelo M Wanderley and Philippe Depalle. Gestural control of sound synthesis. *Proceedings of the IEEE*, 92(4):632–644, 2004.

[174] MM Wanderley and N Orio. Evaluation of input devices for musical expression: Borrowing tools from hci. *Computer Music Journal*, 26(3):62–76, 2002.

[175] Jian-Heng Wang, Siang-An Wang, Wen-Chieh Chen, Ken-Ning Chang, and Herng-Yow Chen. Real-time pitch training system for violin learners. In *Multimedia and Expo Workshops (ICMEW), 2012 IEEE International Conference on*, pages 163–168. IEEE, 2012.

[176] Richard Ward. 6 things to look for when buying a violin, viola, cello, or bass bow. *Strings*, (120), June 2004.

[177] David Wessel and Matthew Wright. Problems and prospects for intimate musical control of computers. *Computer Music Journal*, 26(3):11–22, 2002.

[178] Paul White. Review: Alesis Air FX, multi-effects processor. *Sound on Sound*, January 2001.

[179] Luwei Yang, Elaine Chew, and Khalid Z Rajab. Vibrato performance style: A case study comparing erhu and violin. In *Proceedings of the International Conference on Computer Music Modeling and Retrieval*, 2013.

[180] Luwei Yang, Elaine Chew, and Khalid Z Rajab. Filter diagonalisation method for music signal analysis: Frame-wise vibrato detection and estimation. In *Journal of Mathematics and Music*. Taylor & Francis, 2016. Under Revision.

[181] Luwei Yang, Khalid Z Rajab, and Elaine Chew. Ava: An interactive system for visual and quantitative analyses of vibrato and portamento performance styles. In *Proceedings of the 17th Conference of the International Society for Music Information Retrieval (ISMIR)*, pages 108–114, 2016.

[182] Matthew Yee-King, Mark d'Inverno, et al. Pedagogical agents for social music learning in crowd-based socio-cognitive systems. In *Crowd Intelligence: Foundations, Methods, and Practices (CEUR Workshop Proceedings)*, volume 1148, pages 76–93, 2014.

[183] Matthew Yee-King, Maria Krivenski, Harry Brenton, Andreu Grimalt-Reynes, and Mark d'Inverno. Designing educational social machines for effective feedback. *International Association for Development of the Information Society*, 2014.

[184] Diana Young. The hyperbow: a precision violin interface. In *Proceedings of the 2002 International Computer Music Conference*, 2002.

[185] Diana Young. The hyperbow controller: Real-time dynamics measurement of violin performance. In *Proceedings of the 2002 Conference on New Interfaces for Musical Expression*, pages 1–6, 2002.

[186] Diana Young. Wireless sensor system for measurement of violin bowing parameters. In *Proceedings of the 2003 Stockholm Music Acoustics Conference, SMAC*, pages 111–114, 2003.

[187] Diana Young and Anagha Deshmane. Bowstroke database: a web-accessible archive of violin bowing data. In *Proceedings of the 2007 Conference on New Interfaces for Musical Expression*, pages 352–357. ACM, 2007.

[188] Diana S Young. *New frontiers of expression through real-time dynamics measurement of violin bows*. PhD thesis, Massachusetts Institute of Technology, 2001.

[189] Victor Zappi and Andrew McPherson. Dimensionality and appropriation in digital musical instrument design. In *NIME*, pages 455–460, 2014.

[190] Robert J Zatorre, Joyce L Chen, and Virginia B Penhune. When the brain plays music: auditory–motor interactions in music perception and production. *Nature reviews neuroscience*, 8(7):547–558, 2007.

Appendices

Appendix A

Reference Scores

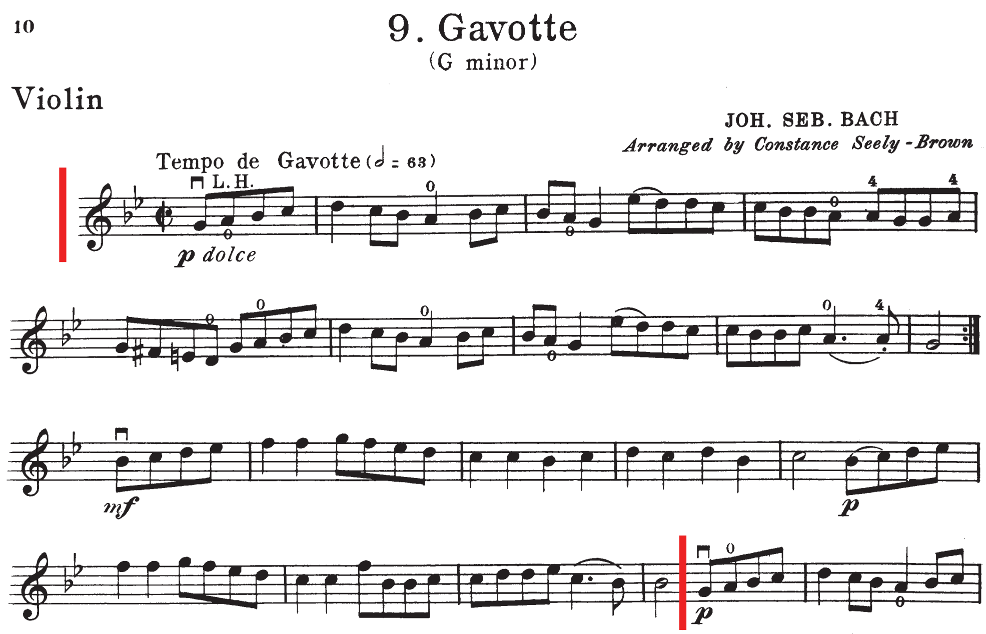


Figure A.1: Score selection from Bach's Gavotte in G minor from BWV 822 (arr. Seely-Brown). The two red bars mark the start and end of the selection used in the study of pitch snap with experts (Chapter 8).

Figure A.2: Score to Bach's Minuet, BWV Anhang 114 (arr. Seely-Brown), also referred to as Minuet 3 in Suzuki repertoire. The two red bars mark the start and end of the selection, bars 17-32, used in the study of pitch snap with experts (Chapter 8). The opening of the piece, bars 1-16, were used for determining pitch accuracy for the common repertoire piece when played in the study with beginners (Chapter 9).

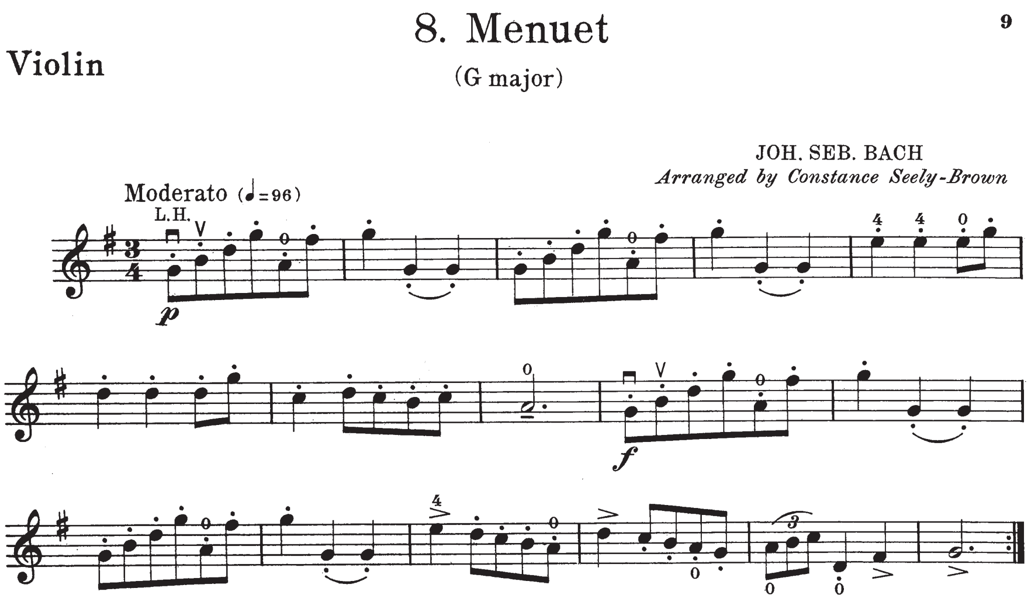


Figure A.3: Score selection from Bach's Minuet in G Major, BWV Anhang 116 (arr. Seely-Brown). This selection was one of the excerpts used in the study of pitch snap with experts (Chapter 8).

Long Long Ago

Figure A.4: Score to T.H. Bayly's Long Long Ago. This piece was used as one of the common repertoire pieces in the study of pitch aids with beginners (Chapter 9).

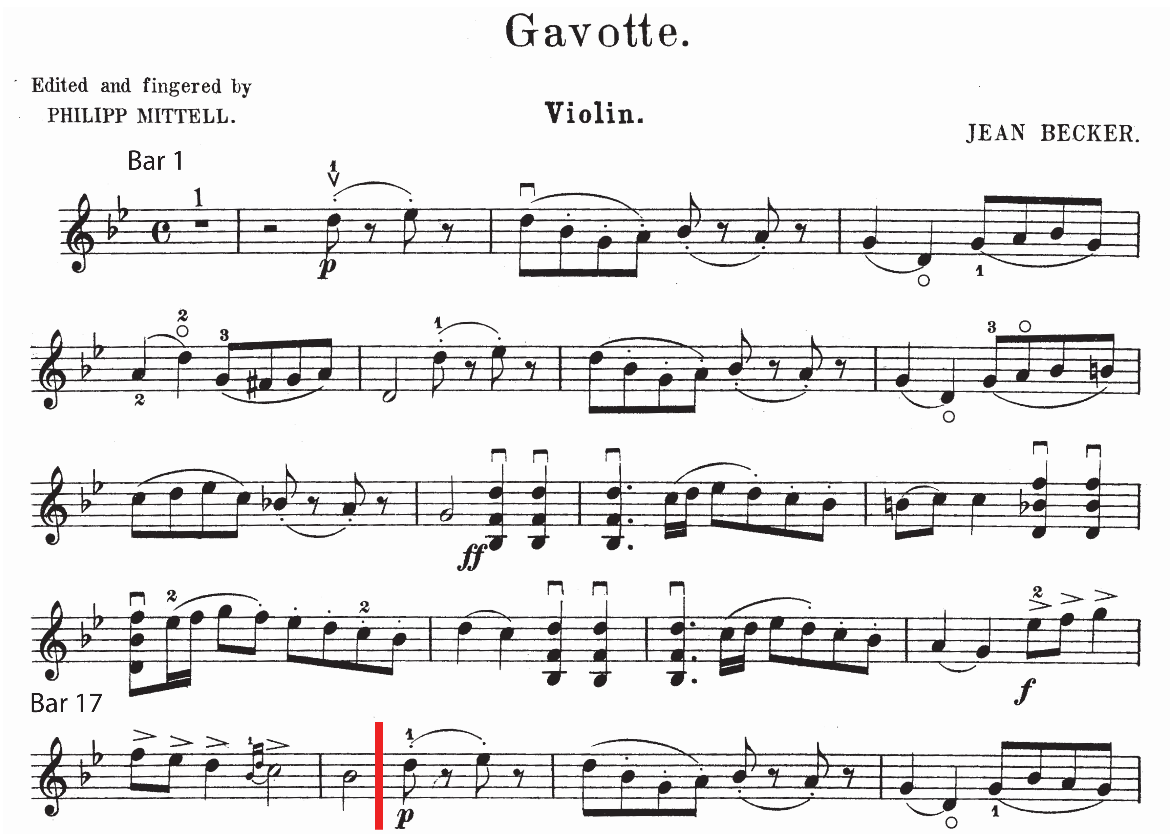


Figure A.5: Score selection from Becker's Gavotte (arr. Moffat). This selection was one of the excerpts used in the study of pitch snap with experts (Chapter 8), though the study used a different arrangement that did not include double stops. The red marker in bar 18 marks the end of the selection.

Figure A.6: Score selection from Beethoven's Minuet In G. The red bar marks the end of the section used in the study of pitch snap with experts (Chapter 8). Participants were given a different edition which only featured the top line and did not include any double stops.

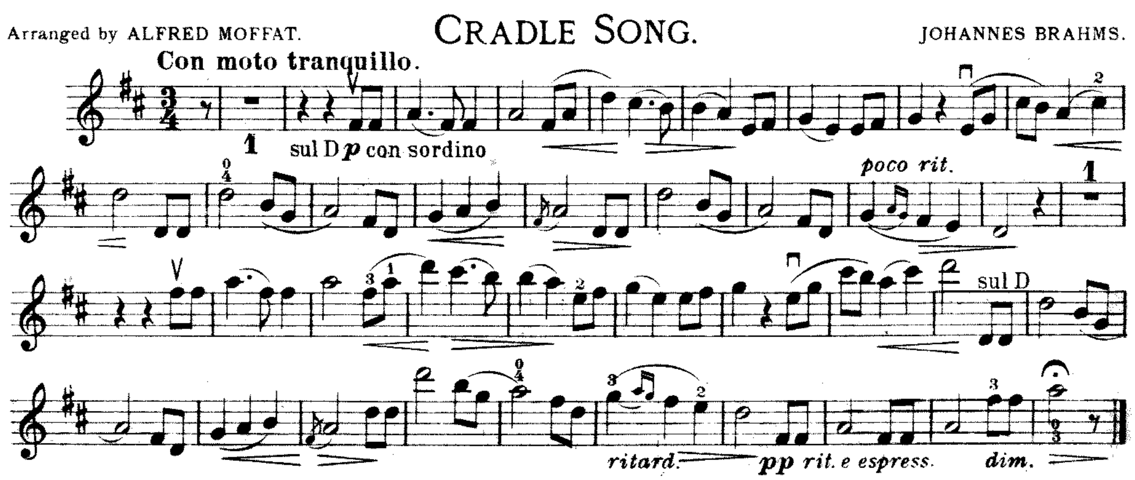


Figure A.7: Score to Brahm's Lullaby or Craddle Song (arr. Mitchel). This piece was one of the excerpts used in the study of pitch snap with experts (Chapter 8).

Figure A.8: Score selection from an arrangement of Brahm's Hungarian Dance No. 5. The red bar marks the end of the section used in the study of pitch snap with experts (Chapter 8). This arrangement is different from the one used in the study which was transposed to D minor.

Estrellita

M. Ponce

Figure A.9: Score selection from M. Ponce's Estrellita (arr. Halle). This piece was one of the excerpts used in the study of pitch snap with experts (Chapter 8).

Serenade

Franz Schubert

Figure A.10: Score selection from F. Schubert's Serenade (arr. Halle). This piece was one of the excerpts used in the study of pitch snap with experts (Chapter 8).

Figure A.11: Score selection from Schumann's Traumerei (arr. Ernst). The red bar marks the end of the selection used in the study of pitch snap with experts (Chapter 8). The selection includes the repeat though this arrangement is different from the one used in the study.

Figure A.12: Score selection from Seitz's Student Concerto No. 5, 1st Movement. The two red bars mark the start and end of the selection used in the study of note onset (Section 6).

Figure A.13: Score selection from Seitz's Student Concerto No. 5, 3rd Movement. The two red bars mark the start and end of the selection used in the study of pitch snap with experts (Chapter 8).

Figure A.14: Score selection from an arrangement of Tschaikowsky's March Slave (arr. Halle). This selection was one of the excerpts used in the study of pitch snap with experts (Chapter 8).

Appendix B

Parent Questionnaire

Questionnaire For Parents About Students:

Please answer to the best of your knowledge. (If you are an adult participant, please answer these same questions about yourself.)

1. Student's Name:

2. Age:

3. How long has your child been learning violin?

4. What is their most recent piece?

5. How much on average does your child practice each week?

6. Does your child learn best through visual learning tools, hearing or being instructed, or through experience? Or are you not aware of any preference?

7. Has your child ever used a tuner at home to help them work on intonation? If yes, how often?

8. Does your child listen to the songs they are learning? How often?

9. Has your child ever used a piano or other intonation aid to help learn the correct notes in a song? Has you child ever played along with a recording during practice? If yes, how often?

10. Does your child's violin have stickers or tapes on the fingerboard to help them know where to place the finger?

Appendix C

Consent Forms

Pro forma information sheet and consent form

Queen Mary
University of London

Information sheet

Research Study, Use of An Augmented Violin to Provide Assistive Feedback to Beginning Violin Students: Information for Participants *Teaching*

We would like to invite you to be part of this research project, if you would like to. You should only agree to take part if you want to, it is entirely up to you. If you choose not to take part there won't be any disadvantages for you and you will hear no more about it.

Please read the following information carefully before you decide to take part; this will tell you why the research is being done and what you will be asked to do if you take part. Please ask if there is anything that is not clear or if you would like more information.

If you decide to take part you will be asked to sign the attached form to say that you agree. You are still free to withdraw at any time and without giving a reason.

Details of study:

This project is to study the effectiveness of using an augmented violin to provide visual and audio feedback to enhance the understanding and accuracy of pitch for a beginning violinist. Violin students must learn to play in-tune, requiring accuracy placing the finger in the right place on the violin and being able to hear if the pitch they are playing is right. We have built an augmented violin capable of enabling high speed pitch estimation and correction. Can feedback enabled by the violin be helpful within the lesson environment to help learn "correct pitch" faster?

We are testing whether providing a student external pitch feedback guides them towards "correct pitch" and whether the feedback helps achieve pitch accuracy faster and enhances the student's perception of "correct pitch". During lessons comprising this study, the student will be given an augmented violin built using a violin and custom sensors that analyses pitch played and gives either aural or visual feedback to help guide them to the correct pitch. Sensors on the bow are used to track its movement, allowing us to investigate whether automatic pitch correction has any secondary effects on control of the bow.

Aural feedback is generated by auto-tuning (pitch shifting) what the student plays to the closest pitch in the scale. The auto-tuned version is played through a speaker. Visual feedback is generated by giving the student a visual with the note name of the nearest pitch (A, Ab, B, etc…) next to a bar scaled in size and coloured depending on how far the student is from the nearest pitch in the scale. The student may also be assigned to use the augmented violin with no additional feedback beyond the sound of the violin itself.

As a teacher, we ask you to nominate students who may be appropriate for joining the study. We also ask you to help assess participating students and their experience with the augmented violin at the beginning and end of the study. We may also ask you for your own thoughts on the augmented violin and the feedback it provides to the student. When communicating about participation with a student, please ensure that it is clear that participation is voluntary and that non-participation will not have any negative repercussions. During the study, all teaching will be in line with normal teaching practice using agreed repertoire.

Lessons will be recorded on video for research analysis purposes only. We will also record audio and performance data from the augmented violin. We will not use or release any video in public, nor any audio which identifies the teacher or student (e.g. conversations), without the explicit written consent of you and the student or their parents if minors. We may use audio clips of violin playing in presenting our research, in which case this will be completely anonymous. All media and data collected as part of this project will be stored securely and anonymously, and individual participants will not be identified in any publications which might arise from this research.

It is up to you to decide whether or not to take part. If you do decide to take part you will be given this information sheet to keep and be asked to sign a consent form.

If you have any questions or concerns about the manner in which the study was conducted please, in the first instance, contact the researcher responsible for the study. If this is unsuccessful, or not appropriate, please contact the Secretary at the Queen Mary Ethics of Research Committee, Room W117, Queen’s Building, Mile End Campus, Mile End Road, London or research-ethics@qmul.ac.uk.

Pro forma information sheet and consent form

Queen Mary
University of London

Information sheet

Research study, Use of An Augmented Violin to Provide Assistive Feedback to Beginning Violin Students: Information for *Parents*

We would like to invite your child to be part of this research project, if you and your child would like to. Your child should only agree to take part if they want to, it is entirely up to you and your child. If you and your child choose not to take part there will not be any disadvantages for you or your child and you will hear no more about it. Choosing not to take part will not negatively affect your child's access to their regular lessons in any way.

Please read the following information carefully before you decide to allow your child to take part; this will tell you why the research is being done and what they will be asked to do if they take part. Please ask if there is anything that is not clear or if you would like more information.

If you decide for your child to take part you will be asked to sign the attached form to say that you agree. You are still free to withdraw at any time and without giving a reason.

Details of study:

This project is to study the effectiveness of using an augmented violin to provide visual and audio feedback to enhance the understanding and accuracy of pitch for a beginning violinist. Violin students must learn to play in-tune, requiring accuracy placing the finger in the right place on the violin and being able to hear if the pitch they are playing is right. We have built an augmented violin capable of enabling high speed pitch estimation and correction. Can feedback enabled by the violin be helpful within the lesson environment to help learn "correct pitch" faster?

We are testing whether providing a student external pitch feedback guides them towards "correct pitch" and whether the feedback helps achieve pitch accuracy faster and enhances the student's perception of "correct pitch". During lessons comprising this study, the student will be given an augmented violin built using a violin and custom sensors that analyses pitch played and gives either aural or visual feedback to help guide them to the correct pitch. Sensors on the bow are used to track its movement, allowing us to investigate whether automatic pitch correction has any secondary effects on control of the bow.

Aural feedback is generated by auto-tuning (pitch shifting) what the student plays to the closest pitch in the scale. The auto-tuned version is played through a speaker. Visual feedback is generated by giving the student a visual with the note name of the nearest pitch (A, Ab, B, etc…) next to a bar scaled in size and coloured depending on how far the student is from the nearest pitch in the scale. The student may also be assigned to use the augmented violin with no additional feedback beyond the sound of the violin itself.

We have worked with your child's teacher to ensure that the study will not interfere with your child's usual lessons. While we ask the student uses and tells us what they think of the augmented violin, bow, and pitch feedback during the duration of the lesson, they will not otherwise be asked to do anything significantly different from what they might normally encounter in their normal lesson.

Lessons will be recorded on video for research analysis purposes only. We will also record audio and performance data from the augmented violin. We will not use or release any video in public, nor any audio which identifies the teacher or student (e.g. conversations), without the explicit written consent of you and your child's teacher. We may use audio clips of violin playing in presenting our research, in which case this will be completely anonymous. All media and data collected as part of this project will be stored securely and anonymously, and individual participants will not be identified in any publications which might arise from this research.

It is up to you to decide whether or not your child is allowed to take part. If you do decide they are allowed to take part you will be given this information sheet to keep and be asked to sign a consent form.

If you have any questions or concerns about the manner in which the study was conducted please, in the first instance, contact the researcher responsible for the study. If this is unsuccessful, or not appropriate, please contact the Secretary at the Queen Mary Ethics of Research Committee, Room W117, Queen's Building, Mile End Campus, Mile End Road, London or research-ethics@qmul.ac.uk.

Pro forma information sheet and consent form

Queen Mary
University of London

Information sheet

Research study, Use of An Augmented Violin to Provide Assistive Feedback to Beginning Violin Students: Information for *Adult Student Participants*

We would like to invite you to be part of this research project, if you would like to. You should only agree to take part if you want to, it is entirely up to you. If you choose not to take part there won't be any disadvantages for you and you will hear no more about it. Choosing not to take part will not negatively affect your access to your usual lessons in any way.

Please read the following information carefully before you decide to take part; this will tell you why the research is being done and what you will be asked to do if you take part. Please ask if there is anything that is not clear or if you would like more information.

If you decide to take part you will be asked to sign the attached form to say that you agree. You are still free to withdraw at any time and without giving a reason.

Details of Study:

This project is to study the effectiveness of using an augmented violin to provide visual and audio feedback to enhance the understanding and accuracy of pitch for a beginning violinist. Violin students must learn to play in-tune, requiring accuracy placing the finger in the right place on the violin and being able to hear if the pitch they are playing is right. We have built an augmented violin capable of enabling high speed pitch estimation and correction. Can feedback enabled by the violin be helpful within the lesson environment to help learn "correct pitch" faster?

We are testing whether providing a student external pitch feedback guides them towards "correct pitch" and whether the feedback helps achieve pitch accuracy faster and enhances the student's perception of "correct pitch". During lessons comprising this study, the student will be given an augmented violin built using a violin and custom sensors that analyses pitch played and gives either aural or visual feedback to help guide them to the correct pitch. Sensors on the bow are used to track its movement, allowing us to investigate whether automatic pitch correction has any secondary effects on control of the bow.

Aural feedback is generated by auto-tuning (pitch shifting) what the student plays to the closest pitch in the scale. The auto-tuned version is played through a speaker. Visual feedback is generated by giving the student a visual with the note name of the nearest pitch (A, Ab, B, etc…) next to a bar scaled in size and coloured depending on how far the student is from the nearest pitch in the scale. The student may also be assigned to use the augmented violin with no additional feedback beyond the sound of the violin itself.

We have worked with your teacher to ensure that the study will not interfere with your usual lessons. While we ask you to use and tell us what you think of the augmented violin, bow, and pitch feedback during the duration of the lesson, you will not otherwise be asked to do anything significantly different from what you might normally encounter in a normal lesson.

Lessons will be recorded on video for research analysis purposes only. We will also record audio and performance data from the augmented violin. We will not use or release any video in public, nor any audio which identifies the teacher or student (e.g. conversations), without the explicit written consent of you and your teacher. We may use audio clips of violin playing in presenting our research, in which case this will be completely anonymous. All media and data collected as part of this project will be stored securely and anonymously, and individual participants will **not** be identified in any publications which might arise from this research.

It is up to you to decide whether or not to take part. If you do decide to take part you will be given this information sheet to keep and be asked to sign a consent form.

If you have any questions or concerns about the manner in which the study was conducted please, in the first instance, contact the researcher responsible for the study. If this is unsuccessful, or not appropriate, please contact the Secretary at the Queen Mary Ethics of Research Committee, Room W117, Queen's Building, Mile End Campus, Mile End Road, London or research-ethics@qmul.ac.uk.

Pro forma information sheet and consent form

Queen Mary
University of London

Information sheet

Research study, Use of An Augmented Violin to Provide Assistive Feedback to Beginning Violin Students: Information for *Young Student Participants*

We would like to invite you to be part of this research project, if you would like to. You should only agree to take part if you want to, it is entirely up to you. If you choose not to take part there won't be any disadvantages for you and you will hear no more about it. Choosing not to take part will not negatively affect your access to your usual lessons in any way.

Please read the following information carefully before you decide to take part; this will tell you why the research is being done and what you will be asked to do if you take part. Please ask if there is anything that is not clear or if you would like more information.

If you decide to take part you will be asked to sign the attached form to say that you agree. You are still free to withdraw at any time and without giving a reason.

Details of Study:

This project is to study the use of a violin with special sensors to provide visual and audio feedback in order to improve the understanding and accuracy of pitch for beginning violinists. Violin students must learn to play in-tune, placing the finger in the right place on the violin and being able to hear if the pitch is right. We have built sensors for the violin to help estimate whether the pitch being played is correct and what the correct pitch should sound like.

We are testing whether extra pitch feedback helps you learn better and helps you know if you are in tune, and if you are sharp or flat. During lessons through this study, you will be asked to use a violin and bow with special sensors. Feedback will be provided through a computer that is either visual, similar to a tuner, or that you can hear, similar to if your teacher was playing the same note. In some cases, even though you are using the sensor violin, there will be no feedback and it will be just like a normal lesson.

We have worked with your teacher to ensure that the study will not interfere with your usual lessons. While we will ask you to use and tell us what you think of the sensor violin, bow, and the pitch feedback during the lesson, you will not otherwise be asked to do anything very different from what you might normally encounter in a normal lesson.

Lessons will be recorded on video for research purposes only. We will also record audio and performance data from the sensor violin. We will not use or release any video in public, nor any audio which identifies you or your teacher without the explicit written consent of your parents and your teacher. We may use audio clips of violin playing in presenting our research, in which case it will be completely anonymous. All media and data collected as part of this project will be stored securely and anonymously, and individual participants will **not** be identified in any publications which might arise from this research.

It is up to you to decide whether or not to take part. If you do decide to take part you will be given this information sheet to keep and be asked to sign a consent form.

If you have any questions or concerns about the manner in which the study was conducted please, in the first instance, contact the researcher responsible for the study. If this is unsuccessful, or not appropriate, please contact the Secretary at the Queen Mary Ethics of Research Committee, Room W117, Queen's Building, Mile End Campus, Mile End Road, London or research-ethics@qmul.ac.uk.

Pro forma information sheet and consent form

Queen Mary
University of London

Consent form

Please complete this form after you have read the Information Sheet and/or listened to an explanation about the research.

Title of Study: ____
Queen Mary Ethics of Research Committee Ref: ________________

- Thank you for considering taking part in this research. The person organizing the research must explain the project to you before you agree to take part.
- If you have any questions arising from the Information Sheet or explanation already given to you, please ask the researcher before you decide whether to join in. You will be given a copy of this Consent Form to keep and refer to at any time.
- *I understand that if I decide at any other time during the research that I no longer wish to participate in this project, I can notify the researchers involved and be withdrawn from it immediately.*
- *I consent to the processing of my personal information for the purposes of this research study. I understand that such information will be treated as strictly confidential and handled in accordance with the provisions of the Data Protection Act 1998.*

Participant's Statement:

I ____ agree that the research project named above has been explained to me to my satisfaction and I agree to take part in the study. I have read both the notes written above and the Information Sheet about the project, and understand what the research study involves.

Signed: Date:

Investigator's Statement:

I ____ confirm that I have carefully explained the nature, demands and any foreseeable risks (where applicable) of the proposed research to the volunteer.

Pro forma information sheet and consent form

Queen Mary
University of London

Consent form on behalf of minors

Please complete this form after you have read the Information Sheet and/or listened to an explanation about the research.

Title of Study: ____
Queen Mary Ethics of Research Committee Ref: ________________

- Thank you for considering taking part in this research. The person organizing the research must explain the project to you before you agree to take part.
- If you have any questions arising from the Information Sheet or explanation already given to you, please ask the researcher before you decide whether to join in. You will be given a copy of this Consent Form to keep and refer to at any time.
- *I understand that if I decide at any other time during the research that my child or I no longer wish to participate in this project, I can notify the researchers involved and be withdrawn from it immediately.*
- *I consent to the processing of my personal child's information for the purposes of this research study. I understand that such information will be treated as strictly confidential and handled in accordance with the provisions of the Data Protection Act 1998.*

Participant's Statement:

I ____ agree that the research project named above has been explained to me to my satisfaction and I agree for my child to take part in the study. I have read both the notes written above and the Information Sheet about the project, and understand what the research study involves.

Signed: Date:

Investigator's Statement:

I ____ confirm that I have carefully explained the nature, demands and any foreseeable risks (where applicable) of the proposed research to the volunteer.